Я. Б. ЗЕЛЬДОВИЧ, И. М. ЯГЛОМ

ВЫСШАЯ МАТЕМАТИКА для начинающих физиков и техников



АКАДЕМИЯ НАУК СССР
ИНСТИТУТ ТЕОРЕТИЧЕСКОЙ ФИЗИКИ ИМЕНИ Л. Д. ЛАНДАУ

Я. Б. ЗЕЛЬДОВИЧ, И. М. ЯГЛОМ

ВЫСШАЯ МАТЕМАТИКА для начинающих физиков и техников



ИЗДАТЕЛЬСТВО «НАУКА»
МОСКВА
1982

УДК 517(07)

Зельдович Я. Б., Яглом И. М. Высшая математика для начинающих физиков и техников. М.: Наука, 1982. 512 с.

Настоящая книга представляет собой введение в математический анализ. Наряду с изложением начал аналитической геометрии и математического анализа (дифференциального и интегрального исчисления) книга содержит понятия о степенных и тригонометрических рядах и о простейших дифференциальных уравнениях, а также затрагивает ряд разделов и тем из физики (механика и теория колебаний, теория электрических цепей, радиоактивный распад, лазеры и др.).

Книга рассчитана на читателей, интересующихся естественнонаучными приложениями высшей математики, преподавателей вузов и втузов, а также будущих физиков и инженеров.

Ил. 363. Табл. 19. Библиогр. 40 назв.

З $\frac{1703040000-341}{055(02)-82}$ 9-81, кн. 2.

ПРЕДИСЛОВИЕ

Эта книга — плод совместного труда физика и математика, поставивших своей целью написать пособие совершенно нового типа для будущих естествоиспытателей. В связи с этим уместно сказать несколько слов о ее целях и научной идеологии.

Известно, что многие физики недовольны тем изложением основ математического анализа, которое было внедрено математиками первой половины XX века. Причины этого понять легко. На стиль изложения даже самых начальных глав учебной литературы для начинающих повлияли итоги развития математических исследований, связанных с разработкой логических основ нашей науки. Точные определения таких понятий, как вещественное число, предел, непрерывность и др., явились результатом длительного и притом весьма нетривиального логического анализа уже созданных и интуитивно ясных (на естественнонаучном уровне строгости) теорий. Эти определения, которые для начинающих вовсе не просты, впоследствии стали использоваться нерационально: в учебниках они стали предшествовать изложению содержательной теории и ее приложений, искусственно усложняя понимание интуитивно ясных вещей.

Следует иметь в виду, что физики-теоретики в своей научной работе широко пользуются всеми или почти всеми методами классической математики. Часто им приходится использовать самые современные идеи, зародившиеся в глубинах абстрактной математики, или даже самостоятельно изобретать новые математические методы. Однако конечный результат исследований физика всегда должен быть эффективен: он должен выражаться числом или формулой, относящимися к наблюдаемым величинам. Вопросы обоснования и логической структуры основ кажутся ему неважными для его работы, хотя он, конечно, верит математике и не сомневается, что в ее методах не заложены никакие внутренние противоречия. Не удовлетворяясь существующей учебной и научно-популярной литературой по математике, написанной самими математиками (в особенности теми, кто далек от физики), физики-теоретики считают необходимым довести до начинающего свои представления о том, как нужно естествоиспытателю пользоваться математическим аппаратом и каким простейшим способом можно «в первом приближении» освоить те методы, которые ему прежде всего понадобятся. Мне кажется, что их громадный опыт работы с математикой дает им на это право.

Конечно, такая точка зрения на математику является несколько односторонней. Но беды в этом, по-моему, нет. Те, кто в дальнейшем будут изучать математику более глубоко, смогут, уже имея интуитивное представление о сути дела и владея техникой решения задач, лучше понять смысл и оснований математики. На мой взгляд, желание начать обучение еще не освоенной теории, которую учащийся познает впервые, с ознакомления с основаниями и формальным языком портит многие математические курсы. Как математик, много лет работавший с физиками, я позволю себе высказать мнение, что не только будущим физикам и техникам, но и будущим профессиональным математикам (и их педагогам) в процессе образования

было бы полезно научиться лучше представлять себе стиль математического мышления естествоиспытателей и понимать стоящие перед ними задачи.

Настоящая книга возникла в результате большой работы над усовершенствованием книги «Высшая математика для начинающих и ее приложения к физике» известного советского физика академика Я. Б. Зельдовича. Цель книги состоит в том, чтобы дать возможность будущему физику (химику, инженеру и др.) использовать в своей работе высшую математику, освоив ее методы без полных логических обоснований, рассматривая ее как раздел естествознания и решая возможно большее число конкретных задач.

Предшествующая книга была не лишена ряда технических дефектов. При написании этой книги много было сделано для устранения недостатков и улучшения изложения; был добавлен также значительный новый материал и проведено тщательное редактирование текста. Вся эта работа была проведена совместно двумя авторами настоящей книги: физиком академиком Я. Б. Зельдовичем и математиком доктором физ.-мат. наук И. М. Ягломом.

Мне кажется, что книга хорошо решает поставленную задачу; можно уверенно рекомендовать ее вдумчивому юному читателю, который хочет неформально овладеть средствами высшей математики и попробовать применять эти средства в задачах физики и техники.

Академик С. П. Новиков

К ЧИТАТЕЛЮ

В нашей стране вряд ли найдется много людей, не знающих Пушкина и Льва Толстого. Однако, вероятно, гораздо большее число взрослых людей — половина или девять десятых — не ответят на вопрос, что такое *производная* и *интеграл*. Между тем без этих понятий невозможно описывать и исследовать переменные величины и функции, характеризующие зависимости одних величин от других. Законы природы формулируются на языке высшей математики, на языке производных и интегралов.

На наш небеспристрастный взгляд, высшая математика так же красива, как стихи Пушкина, и так же глубока, как проза Достоевского или Толстого. Многолетний спор о том, надо ли преподавать основы математического анализа в средней школе, ныне решен Министерством просвещения СССР. Но стало ли общепринятым, вошло ли действительно в общественное сознание отношение к высшей математике как к культурной ценности? Сегодня, когда современная наука далеко раздвинула рамки видимого мира, раскрыв как тайны строения атома, так и многие загадки звездного неба, понятия производной и интеграла стали необходимыми элементами общей культуры. Да и в повседневной жизни представления о скорости изменения той или иной величины (производная) и о суммарном эффекте действия какого-либо фактора (интеграл) достаточно полезны: они расширяют кругозор и применимы в многочисленных и разнообразных ситуациях.

Старые традиции в преподавании высшей математики осложняли дело. Казалось, что легко понять «Преступление и наказание» или «Войну и мир», легко освоиться с шарообразностью Земли или с атомной структурой материи, но трудно овладеть дифференцированием и интегрированием. Учащихся первой половины XX века отпугивала теория пределов, язык «бесконечно малых», столь непривычный после обычной арифметики и школьной алгебры. При традиционном введении высшей математики на первом курсе высших учебных заведений именно математика становилась главной причиной отсева учащихся. Только изменив способ преподавания, можно изменить и отношение к предмету. Высшая математика должна превратиться из сухого и трудного предмета в комплекс ясных и естественных представлений, открывающих прямой путь к изучению физики, химии, инженерно-технических дисциплин.

Итак, первая задача, стоявшая перед авторами при написании настоящей книги, — дать читателю доступное введение в высшую математику, не отягощенное ни громоздким аппаратом, ни логическими тонкостями. Мы убеждены, что каждому, кто знаком с азами арифметики и имел по алгебре в 6-м и 7-м классах средней школы отметки, в среднем превышающие «три», будет нетрудно (и, мы надеемся, интересно) читать нашу книгу. Мы рассчитываем на читателя-друга, который хочет не сомневаться, а верить и берется за книгу для того, чтобы научиться новому, не смущаясь тем, что «высокоученую» теорию авторы иногда заменяют просто разбором примеров.

В книге рассмотрено много конкретных примеров (для кого-то, быть может, даже и слишком много). Здесь и численные расчеты, и задачи, связанные с теми или иными явлениями природы. Мы надеемся, что эти примеры облегчат читателю изучение книги. Заметим, наконец, что, имея дело с читателем — завзятым спорщиком, мы писали бы по-другому. Такому читателю, быть может, больше подходит стандартное изложение со всеми принятыми в таких случаях оговорками, но мы не хотим из-за одного спорщика терять 100 возможных друзей.

Эта книга рассчитана на начинающих — на школьников-старшеклассников, учащихся ПТУ и техникумов, студентов младших курсов вузов. Думая о ее читателях, мы видим перед собой также и тех, кто пожелает самостоя-

тельно познакомиться с высшей математикой, скажем людей, окончивших школу в те безвозвратно канувшие в прошлое, но так еще близкие к нам времена, когда дифференциальное и интегральное исчисление в школе не проходилось.

Место математики в жизни и в науке определяется тем, что она позволяет перевести «общежитейские», интуитивные подходы к действительности, базирующиеся на чисто качественных, а значит, приблизительных описаниях, на язык точных определений и формул, из которых возможны количественные выводы. И не случайно говорят, что степень научности той или иной дисциплины измеряется тем, насколько в ней применяется математика. Однако истинным языком науки являются вовсе не элементарная алгебра и геометрия. Гораздо большую роль играет высшая математика, изучающая переменные величины и текущие процессы, для анализа которых алгебры и геометрии недостаточно (а также другие более сложные разделы математики, которых мы в этой книге лишь бегло коснемся). И не случайно само возникновение дифференциального и интегрального исчисления у Ньютона было неразрывно связано с созданием тем же Ньютоном начал теоретической (можно также сказать — математической) физики: понятия высшей математики он воспринимал как (дословный!) перевод на математический язык основных понятий механики [1].

Первые главы книги дают представление как о существе высшей математики, так и о путях ее возможных применений. Таким образом, вполне можно представить себе читателя, который пожелает ограничиться выборочным знакомством с главами 1—7, — но писали мы книгу не для него. Овладение основами высшей математики неразрывно связано с серьезным пониманием путей применения этого аппарата. Знания без применений и настоящими-то знаниями не являются — они с трудом усваиваются, да и забываются легко. Наше пособие рассчитано на лиц, специально интересующихся физикой и техникой; поэтому и приложения здесь в основном заимствованы из этих областей. Сегодня можно было бы составить учебник высшей математики, ориентированный и на биолога (или будущего биолога), геолога или экономиста, однако это были бы совсем другие книги и писать их должны другие.

Вторую часть книги составляют изложения ряда тем из физики и техники. Вот здесь целиком проявляется сила высшей математики: весьма многие важные физические явления удается описать с той степенью законченности, которая недостижима без использования понятий производной и интеграла. В качестве примера мы можем назвать теорию *радиоактивности* (гл. 8). Другим примером может служить явление *резонанса*, выступающее в книге в двух разных обличиях: в виде одного из эффектов, возникающих в теории механических колебаний (гл. 10), и как одно из свойств электрических цепей (гл. 13). Несмотря на то, что одни из разделов физики изложены во второй части книги с большой степенью подробности, а другие только намечены или даже вовсе опущены, общий объем даваемых здесь сведений по физике настолько велик, что книгу в целом можно рассматривать не только как пособие по математике, но и как важное добавление к традиционным учебникам физики.

Посвященные физическим вопросам главы 8—13 (они составляют вторую часть книги) почти независимы одна от другой; читать их можно в любом порядке. Единственное исключение здесь составляет глава 10 («Колебания»), для понимания которой нужны первые шесть параграфов предшествующей главы 9 («Механика»). Заметим, что хотя и первая часть книги («Элементы высшей математики») богата примерами из физики и механики, озна-

[1] Столь же неслучайно наличие у второго творца высшей математики — Лейбница эволюционных представлений, тесно связанных с его математическими концепциями и значительно перегнавших свое время. (Эти представления затрагивали биологию, геологию, физику: так, от Лейбница ведет начало концепция энергии, лишь трансформирующейся, но не исчезающей ни при каких физических явлениях.)

комление с какими-то разделами второй части («Приложения высшей математики») полезно, как нам кажется, для всех без исключения категорий читателей.

Большой объем настоящей книги, казалось бы, противоречит слову «начинающие» в ее заголовке. Пусть Вас не смущает количество страниц в этом пособии! Мы хорошо сознаем все недостатки, — но также и достоинства — принятой нами подробной, откровенно многословной системы изложения. Книга содержит очень много примеров и численных расчетов: одни и те же вопросы освещаются в ней зачастую по несколько раз и притом с разных точек зрения; много места уделено истории излагаемых теорий и т. д. Разумеется, всю имеющуюся в книге информацию можно было бы «спрессовать» и изложить на гораздо меньшем числе страниц, но тогда чтение книги требовало бы больших усилий и любое пропущенное или неверно понятое место затрудняло бы (если не делало невозможным) дальнейшее продвижение. Конечно, в жизни встречаются ситуации, когда лучшим источником информации служит краткая телеграмма, но столь любимое XVIII и XIX столетиями и почти забытое в наши дни искусство неторопливого письма, не пренебрегающего многочисленными, казалось бы необязательными, подробностями, также имеет свою прелесть. И для начинающего «телеграфный» стиль, как нам кажется, никак не подходит: избыточная информация, которой в отдельных случаях можно и пренебречь, на плечи не давит, а вот недостаток информации или чрезмерная ее концентрация в книге для неопытного читателя совершенно недопустимы!

Необходимо также иметь в виду, что наша книга содержит много материала для лиц, заинтересованных в углубленном изучении основ высшей математики; при первом чтении значительную часть этого материала можно и пропустить. Сюда относятся, например, все разделы книги, напечатанные мелким шрифтом (в частности, рассказ о происхождении высшей математики) или отмеченные звездочками параграфы. Подобный характер имеет и вся третья часть книги («Дополнительные темы из высшей математики») и Заключение («Что же дальше?»). Наконец, мы уже подчеркивали возможность выборочного знакомства со второй (физической) частью книги (гл. 8—13).

Большая глава 1 «Функции и графики» в основном содержит материал, известный из средних классов школы; многим читателям будет достаточно бегло ее просмотреть. Главы 2 и 3 излагают идейную сторону высшей математики (дифференциального и интегрального исчисления), а главы 4 и 5 — аппаратную ее сторону в том достаточно скромном объеме, который необходим для основных применений. Глава 6 посвящена некоторым более продвинутым разделам математического анализа (ряды, дифференциальные уравнения), а глава 7 касается чисто математических применений понятий производной и интеграла (физические применения рассматриваются во второй части книги). Элементы истории высшей математики отнесены в заключительные разделы глав 3 и 6, отделенные от основного текста тремя звездочками. К математической части книги относится и Заключение «Что же дальше?», посвященное обсуждению некоторых более новых разделов математической науки, ныне широко используемых в приложениях.

Укажем, что третья часть книги (и особенно ее Заключение) имеют несколько иной характер, чем первая и вторая части. Последняя часть книги посвящена двум направлениям высшей математики, существенно расширяющим (и обобщающим) классическое дифференциальное и интегральное исчисление Ньютона и Лейбница. Первое из них — *комплексные числа* и *функции комплексной переменной*, которые давно уже играют в физике и технике громадную роль. Замечательно, в частности, что многие обстоятельства, касающиеся обычных функций с вещественными аргументами и значениями, становятся яснее при обращении к комплексным числам. Некоторые примеры такого рода читатель найдет в главах 14, 15 и 17.

Второе направление, рассматриваемое в третьей части книги, — так называемые *обобщенные функции* $f(x)$, в первую очередь дельта-функция Дирака, равная нулю всюду, кроме значения $x=0$, при котором она равна бесконечности (!). Такие функции также играют в физике (и в технике) большую роль; само их происхождение гораздо больше связано с физикой, чем с математикой. Нам представляется очень желательным возможно более ранее ознакомление будущих физиков и инженеров с этими замечательными функциями!

Посвященные функциям комплексной переменной и обобщенным функциям главы 14—17 рассчитаны лишь на особо интересующихся читателей. Написаны они несколько более конспективно, чем остальной текст книги (хотя и они доступны достаточно настойчивому начинающему читателю); их можно просмотреть с любой степенью подробности.

Еще более обзорный характер имеет Заключение «Что же дальше?». Вот здесь мы никак не рассчитываем на полное понимание материала, призванного лишь пробудить любознательность читателя и побудить его к дальнейшей работе. Наивно было бы думать, что математическое образование физика или инженера может завершиться темами, охваченными нашей книгой. В Заключении мы коснулись некоторых иных направлений науки, начавших широко применяться в физике и технике лишь во второй половине нашего столетия [2]. Заметим, что хотя Заключение (да и вся третья часть книги тоже) свободно могут быть опущены (для понимания основной части книги они не нужны), мы все же рассчитываем, что большинство читателей захотят хоть бегло с ними познакомиться, поскольку они открывают дальнейшую перспективу и знакомят с некоторыми идеями, достаточно важными и характерными для современной науки.

Таким образом, настоящая книга допускает ряд вариантов ее использования. Читателю, заинтересованному в первую очередь в математике, естественно уделить наибольшее внимание первой части книги; однако при этом безусловно будет полезным и ознакомление с теми или иными разделами второй части — скажем с главой 8, или с первыми параграфами главы 9 и частью главы 10, или с главой 13. Тому, кто пожелает еще углубить свои знания по математике, можно порекомендовать обратиться к заключительным параграфам главы 10 или (и) к третьей части книги; при этом читателю, заинтересовавшемуся функциями комплексной переменной и дельта-функциями, никак не следует пренебрегать главой 17, посвященной приложениям соответствующих теорий. Дополнительные сведения по математике можно извлечь из Заключения, однако для серьезного знакомства с затронутыми в Заключении темами безусловно необходимо обратиться к иной литературе. Наконец, читатель, интересующийся более физикой, чем математикой, может довольно бегло ознакомиться с первой частью, полностью игнорируя текст, напечатанный мелким шрифтом и пропуская большинство отмеченных звездочками параграфов (но не § 7 гл. 6, демонстрирующий пути применения дифференциальных уравнений к анализу естественнонаучных явлений). Такому читателю не следует особенно задерживаться на главе 7 первой части, имеющей чисто математический характер; но вот главу 6 он должен проштудировать достаточно внимательно, поскольку ее прикладное значение очень велико. Вторую часть книги (а, может быть, также и третью, тоже связанную с глубокими физическими теориями) интересующийся физикой читатель, видимо, будет изучать особенно подробно; при этом порядок знакомства с главами 8—13 он вполне может выбрать сам — так, например, в ряде отношений целесообразно не разрывать особенно во времени изучение глав 10 и 13.

---

[2] Здесь можно указать прежде всего на теорию групп, которая была развита математиками в прошлом веке и в то время казалась далекой от всякой физики. В настоящее же время успехи физики элементарных частиц в большой степени связаны именно с теорией групп.

В списке литературы, приведенном в конце книги, указаны некоторые другие издания, излагающие основы высшей математики, а также более широкие по охвату материала учебники и книги, по тематике далеко выходящие за рамки изложенного здесь материала. В качестве естественного продолжения настоящего пособия читателю можно порекомендовать книгу [15] [3]. Напротив, книги [19—21] от основного содержания нашей книги довольно далеки: они посвящены ветвям математической науки, примыкающим к направлениям, упомянутым в Заключении.

Из литературы по физике хочется в первую очередь порекомендовать читателю замечательный многотомный лекционный курс [34] (его последний девятый выпуск составляют задачи и упражнения ко всему курсу). Отметим, кроме того, что в книгах [22—30, 37] высшая математика не используется.

В Приложениях I—IV к книге приведены краткие таблицы производных, интегралов и числовых рядов (что это такое, Вы узнаете из первой части книги), а также извлечения из числовых таблиц, придающие всему пособию известную законченность: они позволят читателю при выполнении упражнений, завершающих каждый параграф книги, не обращаться к другим учебникам. В физических примерах мы всюду исходили из принятой в настоящее время международной системы единиц СИ; читателю, недостаточно знакомому с этой системой, может быть полезно Приложение V. Наконец, последнее Приложение VI составляют широко используемые в книге латинский и греческий алфавиты.

В книге принята «дробная» система нумерации формул, упражнений, рисунков: во всех случаях первая цифра (или первое число) указывает номер главы, а вторая — номер параграфа (в рисунках к Заключению или Ответам первые две цифры заменяются буквами З или О). Таким образом, ссылка, скажем, на формулу (10.4.3) указывает на формулу (3) из § 4 главы 10, причем такой характер может иметь ссылка лишь в н е пределов главы 10. Указание на формулу (4.3) отсылает читателя к формуле (3) из § 4 т о й ж е главы; наконец, просто ссылка на формулу (3) означает, что речь идет о третьей формуле т о г о ж е самого параграфа. Та же система принята и в ссылках на параграфы книги, где, например, запись § 10.4 отсылает читателя к § 4 главы 10.

Разумеется, как любое пособие по математике, эта книга предполагает активную работу читателя (по возможности с ручкой или карандашом в руке) — решение приложенных к книге упражнений. Еще лучше, если, кроме того, на столе перед Вами лежит микрокомпьютер: тогда Вы сможете проверить и подкрепить общие теоремы с помощью численных расчетов, сможете свободнее пользоваться методом «арифметического экспериментирования», т. е. приближенными вычислениями, раскрывающими содержание вводимых понятий и смысл формул. Конкретные примеры помогут Вам оценить точность рассыпанных по всей книге приближенных формул, и Вы лучше почувствуете мудрость предков, создававших свои методы и концепции, не обладая современной вычислительной техникой. Одновременно с этим Вы почувствуете и новую силу, которую нам, детям XX века, придают компьютеры.

---

[3] Так, студентам технических вузов, проштудировавшим настоящую книгу на первом году учебы, уместно на втором курсе обратиться к книге [15].

ПРЕДИСЛОВИЕ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

Настоящая книга адресована не только учащимся, но и их преподавателям, знакомым с излагаемым здесь материалом; для них интерес может представлять рекомендуемая нами методика преподавания высшей математики будущим физикам и техникам. Книга не является учебником, рассчитанным на какую-то определенную категорию учащихся: она может быть использована при преподавании математики или физики в средней школе, техникуме или ПТУ, в высшем учебном заведении. Учителю среднего учебного заведения наша книга поможет освежить полученные в вузе представления о связи и взаимозависимости математики и физики. Содержание книги школьный учитель может использовать в факультативных занятиях или в (физических и математических) кружках; кое-что из этой книги может быть рассказано и прямо на уроке. Преподавателю вуза книга может доставить материал для использования в практических занятиях по курсу математического анализа (высшей математики) или общей физики. При этом наше пособие может быть полезно преподавателю, ведущему упражнения по дифференциальному и интегральному исчислению с начинающими студентами технического вуза или физического факультета (отделения) даже в том случае, если на лекциях по курсу математического анализа господствует иной дух. Проработка с учащимися тем, заимствованных из второй (а частично, возможно, и из третьей) части книги, поможет студентам обрести сбалансированную точку зрения на математику и облегчит им усвоение отдельных разделов физики и технических дисциплин, таких, как теория колебаний. Она также — и это, конечно, важнее всего — поможет им не только понять, что именно изучает дифференциальное и интегральное исчисление, но и осознать, как и зачем была создана и зачем нужна эта важнейшая математическая дисциплина.

Список литературы в конце книги рассчитан как на учащихся, так и на преподавателей. Он содержит ряд книг [4—6, 12—14], демонстрирующих возможные пути слитного изложения высшей математики и ее приложений; к сожалению, большинство из них ныне малодоступно даже преподавателю.

Книга состоит из трех частей, первая из которых (главы 1—7) в основном напоминает читателю известные ему еще из школьного курса сведения. В этой части излагается идейная и аппаратная сторона высшей математики. Вторая часть (главы 8—13) посвящена приложениям высшей математики к физике и технике; она составляет органическую часть всего сочинения, в то время как третья часть (главы 14—17) является дополнительной и в принципе может быть опущена. Неразрывная связь первой (математической) и второй (физической) частей книги еще подчеркивается тем, что, скажем, раздел о рядах Фурье отнесен во вторую часть пособия (в главу о колебаниях), а изучение процесса вытекания воды составляет часть главы 6, посвященной рядам и дифференциальным уравнениям.

При составлении книги мы не ориентировались специально на современное положение дел, когда учащиеся могут пользоваться микрокомпьютерами. Разумеется, наличие микрокомпьютеров может еще упростить изложение. Оно, в частности, делает гораздо более легкими и быстрыми (а значит, и более естественными) разного рода «арифметические эксперименты», вроде фигурирующего в § 4.7, и позволяет гораздо шире их использовать. Наличие микрокомпьютеров позволяет также гораздо шире применять численное (приближенное) дифференцирование и интегрирование; его естественно принимать во внимание и при изложении приложений математики.

Настоящая книга связана с другой книгой одного из ее авторов, в первоначальном своем виде вышедшей в свет в 1960 г.[1] Одним из стимулов к ее

[1] Ее последнее издание: *Зельдович Я. Б.* Высшая математика для начинающих и ее приложения к физике. 5-е изд. М.: Наука, 1970.

составлению явилось то, что в те годы в средних школах дифференциальное и интегральное исчисления даже не упоминались: в своей аналитической части школьный курс математики ограничивался началами алгебры и тригонометрией, все еще сохранявшей положение самостоятельной дисциплины. В вузовском же преподавании господствовала система изложения начал анализа, весьма далекая от интересов физиков и техников. Напомним, что все 30-е годы основным втузовским учебником высшей математики в нашей стране являлась известная книга В. Гренвиля, переведенная на русский язык, а впоследствии и переработанная академиком Н. Н. Лузиным; в разных своих вариантах эта простая книга переиздавалась множество раз и была весьма популярна среди учащихся инженерных вузов. Однако ожесточенная критика «чистых» математиков, обвинявших этот учебник в нестрогости, привела к тому, что он был полностью вытеснен из педагогической практики и заменен книгами совсем другой направленности. В 50-е годы изложение математического анализа традиционно начиналось с теории пределов, базировавшейся на отточенной «ε—δ-технике» и оперировавшей глубокими понятиями, ценность которых первоначально лишь декларировалась, но никак не раскрывалась. Все это создавало положение, при котором понятия производной и интеграла, интуитивно достаточно простые и наглядные, представлялись непосвященному чем-то весьма возвышенным и вызывали у него чувства, близкие к мистическому трепету (у студентов-первокурсников к этому еще добавлялся реальный страх получить двойку на экзамене). При этом многие математики, к сожалению, поддерживали такое отношение, всячески подчеркивая логическую тонкость понятия предела и необходимость опереться здесь на какой-либо вариант теории вещественного числа (хорошо известно, однако, что все такие обоснования, по необходимости достаточно сложные, возникли лишь во второй половине XIX века, т. е. двумя столетиями позже создания высшей математики). Логическая требовательность учащихся воспитывалась на теоремах, которые им казались ясными и без всякого доказательства. Так, например, тот факт, что предел у переменной может быть только один [2], традиционно устанавливался без всякого обращения к интуитивному представлению о пределе, с полной достоверностью убеждавшему нас, что если переменная мечется между двумя «пределами», то она не стремится ни к одному.

В настоящее время положение дела, безусловно, изменилось к лучшему. В средних школах проходится обязательный курс начал дифференциального и интегрального исчисления. Также и среди новых пособий для высшей школы имеются стоящие на близких к нам позициях. Так, в выдержавшем ряд изданий учебнике [9] не доказывается ни одна теорема о пределах, ибо автор предпочитает заменять здесь строгую логику откровенной апелляцией к здравому смыслу. Да и те пособия, которые традиционно начинают изложение с развернутой теории пределов, совсем по-другому читаются учащимся, еще из средней школы знакомым с основными понятиями производной и интеграла!

При всем том нам кажется необходимым подробно мотивировать свои методические позиции, тем более что согласны с ними, безусловно, не все. Элементы математического анализа твердо вошли в школьный курс математики, но еще далеко не решен вопрос о том, как они будут там излагаться. Здесь вполне возможен «инженерный», или «естественнонаучный», уровень строгости, но есть и сторонники принципиально иного подхода. Нельзя считать решенным и вопрос о принципах построения курса математики в техникумах и ПТУ, хотя нам кажется, что уж здесь никак не нужен уровень строгости, превышающий принятый в настоящей книге.

---

[2] Ср. с теоремой из современного американского школьного учебника геометрии (*Моиз Э. Э., Даунс Ф. Л.* Геометрия. М.: Просвещение, 1972, с. 55): каждый отрезок имеет середину, и притом только одну, — также явно обращенной к одним лишь будущим математикам.

Ясно, что, прежде чем планировать систему обучения математике, необходимо задуматься, зачем мы ее учим. Ведь ответ на вопрос «как?» невозможен, если нет ответа на вопрос «зачем?». Математику можно учить для того, чтобы:

1) успешно сдать экзамен;
2) развить способность к абстрактному мышлению и логические навыки;
3) овладеть современной вычислительной техникой;
4) применять полученные знания к изучению реальных явлений в технике, физике, биологии, к изучению микромира атомов и элементарных частиц или макромира звезд и галактик.

Все эти цели совершенно разные; они диктуют абсолютно несходные методики и системы обучения. Первая из них, при бесспорной практической значимости, конечно, не должна приниматься во внимание: ясно, что экзаменатор должен приноравливаться к потребностям учащихся, а не учащиеся — к вкусам экзаменатора (хотя реально так бывает, к сожалению, не всегда). Вторая цель не может и не должна полностью игнорироваться: ведь здесь речь идет о подготовке будущих математиков, и рост значения математики заставляет отнестись к этой проблеме с достаточным вниманием. Нелепо лишь считать, что согласованное с целью 2) преподавание должно иметь массовый характер — ведь математики никогда не будут составлять сколько-нибудь заметную часть всего населения. И с этой точки зрения существующий во многих странах уклон школьного курса математики в сторону так называемой современной математики, понимаемой как начала общей теории множеств и символической логики, учение об абстрактно вводимых функциях и так называемых бинарных отношениях, кажется нам вредным. Не может служить основным ориентиром в планировании массового математического образования и цель 3) — вопрос о подготовке будущих программистов и специалистов по ЭВМ является серьезным, но все же достаточно частным.

Наша книга адресована будущим инженерам и естествоиспытателям, ориентирующимся на цель 4) изучения математики. Этой безусловно очень обширной категории учащихся надо, по нашему мнению, скорейшим путем сообщить необходимые для приложений сведения, не загромождая их излишними логическими тонкостями и не стремясь к максимальной общности и полной строгости: ведь для учащихся этого рода математика навсегда останется аппаратом, языком описания тех или иных фактов, а не сущностью, которая интересует их сама по себе. При этом, как нам кажется, и сегодня, как это было в XIX веке и будет, вероятно, в XXI веке, надо в первую очередь делать упор на элементы математического анализа, на дифференциальное и интегральное исчисление и простейшие дифференциальные уравнения.

В подходе к основным понятиям анализа следует исходить из тезиса о том, что природа устроена достаточно просто, что «бог не злонамерен», по известному выражению Эйнштейна. Гипертрофия теорем существования может лишь отпугнуть учащегося, одаренного здоровым физическим чутьем. Ясно, что движущаяся частица имеет скорость, а кривая линия имеет касательную; доказывать существование скорости так же не нужно, как доказывать существование площади плоской фигуры: да вырежьте Вы фигуру из бумаги или картона и взвесьте ее, поделите массу на массу одного квадратного метра материала — вот Вам и площадь! Другими словами, теоремы существования производной и интеграла на первом этапе изучения анализа не нужны (и даже вредны): оба эти понятия имеют естественный физический смысл, а значит, они существуют [3]. Ведь было бы нелепо начинать изучение грамматики не с общих правил, а с исключений или начинать изучение языка не с перво-

---

[3] Когда студенты знаменитого Уильяма Томпсона, лорда Кельвина (1824—1907), пытались определять производную «по Коши», то это только раздражало старого профессора. «Да бросьте вы этого Тотгентнера, — говорил он тогда, — ведь производная — это скорость». (Тотгентнер — профессор «чистой» математики, читавший во времена Кельвина в Кембриджском университете курс математического анализа).

начального запаса слов, а с грамматики. Между тем именно этому подобна, к сожалению, достаточно распространенная система изучения «математического языка», при которой концепция (и свойства) предела предшествует любым содержательным применениям этого понятия.

В биологии существует глубокий закон, утверждающий, что «онтогенез повторяет филогенез», т. е. отдельная особь в своем развитии в той или иной мере повторяет эволюционный путь развития всего рода, к которому эта особь принадлежит: так, зародыш человека в течение определенного периода своего развития имеет жаберные щели, ибо жизнь первоначально зародилась в море и предками млекопитающих (в том числе и человека) были рыбы. Нам кажется, что этот биологический закон следует помнить и педагогам. Будущий физик или инженер должен предварительно изучить математику в таком виде, в каком она создавалась великими учеными XVII и XVIII веков, а не получить ее сразу в том обличии, которое выработалось в результате длительной эволюции ко второй половине XX столетия.

Разумеется, сказанное никак не следует понимать, как призыв к точному воспроизведению в современных учебниках представляющихся сегодня достаточно архаичными по форме рассуждений Ньютона, с его давно забытой терминологией и символикой, или Лейбница, со свойственной последнему «метафизикой бесконечно малых». Однако идейная сторона курса для начинающих вполне может, как нам кажется, строиться на привычных классикам наглядных представлениях, а не на современных теориях вещественного числа и не на сложившихся в течение столетий строгих определениях. Конечно, в какой-то момент учащихся надо проинформировать о возможных осложнениях, но не об экзотических «непрерывных функциях, нигде не имеющих производных» [4], а просто о возможностях разрыва или излома кривой. Однако соответствующие предостережения кажутся нам уместными лишь п о с л е интуитивного освоения основных понятий, а не до этого, подобно тому как ознакомление с каким-либо новым механизмом нелепо н а ч и н а т ь с указаний на возможность его поломки. Кроме того, мы склонны считать гораздо более поучительным (и информативным) не традиционный нигилистический, а современный «конструктивный» подход к особенностям рассматриваемого рода, утверждающий, что каждая (хотя бы и негладкая!) непрерывная функция имеет производную, которая, однако, может оказаться не обычной функцией, а дельта-функцией; этому подходу мы уделяем большое внимание в главе 16 книги. Мы пользуемся также всеми случаями для того, чтобы подчеркнуть такие важные общие идеи как *линейность* и *принцип суперпозиции*, иногда недопустимо оставляемые в тени в книгах, рассчитанных на (настоящих или будущих) физиков и инженеров.

Современная математика развивается весьма бурно, и при этом бурно растет и количество знаний, необходимых ее потребителям. Учить физику или инженеру сегодня приходится больше — некоторые из новых важных для приложений разделов науки названы в Заключении «Что же дальше?». Наряду с этим громадное значение приобретает в наши дни почти не затронутая в этой книге вычислительная математика, базирующаяся на огромных возможностях ЭВМ. Мы считаем вполне возможным более быстрое ознакомление начинающих с первыми разделами курса высшей математики, позволяющее раньше применять свои знания и освобождающее время для изучения новых содержательных математических теорий.

Педагогические позиции авторов многих известных нам учебников математического анализа напоминают установки средневековых схоластов в ди-

---

[4] Здесь мы склонны присоединиться к одному из крупнейших аналитиков второй половины XIX века — французу Шарлю Эрмиту, который писал своему другу голландскому математику Стильтьесу, что он «с ужасом и отвращением отворачивается от этой разрастающейся язвы функций, не имеющих производных». Стоящая на прямо противоположных позициях книга французского математика Б. Мальденброта «Фрактали» (*Mandelbrot B.* Fractals. San Francisco: W. H. Freeman, 1977) очень интересна, но она явно рассчитана на читателей, гораздо более изощренных по сравнению с теми, к кому обращен наш элементарный учебник.

спутах и научных турнирах. Учащийся представляется таким авторам опытным противником, выискивающим слабые места в позиции учителя; задача же педагога сводится к опровержению всех возможных возражений. В противоположность этому мы рассматриваем учащегося как друга и союзника, готового поверить педагогу или учебнику и заинтересованного в первую очередь в том, чтобы побыстрее получить возможность использовать в изучении природы и в технике те новые приемы, которым его научили. Понимание сути нового аппарата, так же как и интуиция, предостерегающая нас в тех случаях, когда аппарат отказывается работать, приходит к такому учащемуся главным образом в результате разбора многих примеров и приложений. При строго логическом подходе вопрос о происхождении, значении и пользе изучаемых понятий и теорем остается в тени. В предлагаемой книге основное внимание уделено «грубым» математическим идеям и связи этих идей с явлениями природы.

Поддержку наших позиций мы находим у многих выдающихся ученых. Знаменитый русский кораблестроитель академик А. Н. Крылов (1863—1945), ученый, техник и инженер с необычайно широким диапазоном интересов и результатов, автор первой на русском языке книги по численным методам математики, переводчик «Математических начал натуральной философии» Ньютона и авторитет в области небесной механики, писал: «Нельзя считать недостаточно строгим для 16-летнего гимназиста то, на чем сам Ньютон основывал все современное учение о мироздании и что он положил в основу своих неопровержимых доказательств строения системы мира» — и указывал, что начинающие зачастую воспринимают утонченную строгость доказательств как «торжество науки над здравым смыслом». Хорошо известна автобиография Альберта Эйнштейна — одного из величайших физиков всех времен, который писал:

«В возрасте 12—16 лет я ознакомился с элементами математики, включая основы дифференциального и интегрального исчисления. При этом, на мое счастье, мне попались книги, в которых обращалось не слишком много внимания на логическую строгость, зато хорошо была выделена всюду главная мысль. Вот это занятие было поистине увлекательно: в нем были взлеты, не уступавшие «чуду» элементарной геометрии, — основная идея аналитической геометрии, бесконечные ряды, понятия дифференциала и интеграла».

Сходной была и позиция замечательного советского физика-теоретика, лауреата Ленинской и Нобелевской премий, академика Льва Давидовича Ландау (1908—1968), который наряду с выдающейся физической интуицией обладал и высокой математической техникой. Однажды Ландау попросили дать отзыв на программу по математике одного из московских учебных институтов физического профиля. Вот что он написал в своем отзыве:

«К сожалению, Ваши программы страдают теми же недостатками, какими обычно страдают программы по математике, превращающие изучение математики физиками наполовину в утомительную трату времени. При всей важности математики для физиков, физики, как известно, нуждаются в считающей аналитической математике; математики же, по непонятной мне причине, подсовывают нам в качестве принудительного ассортимента логические упражнения. . . Мне кажется, что давно пора обучать физиков тому, что они сами считают нужным для себя, а не спасать их души вопреки их собственному желанию. Мне не хочется дискутировать с достойной средневековой схоластики мыслью, что путем изучения ненужных им вещей люди будто бы научаются логически мыслить. Я категорически считаю, что из математики, изучаемой физиками, должны быть полностью изгнаны всякие теоремы существования, слишком строгие доказательства и т. п.»

Эти расхождения между математиками и физиками привели к тому парадоксальному положению, когда трое (!) лауреатов Нобелевской премии по физике и химии (Вальтер Нернст, Сванте Аррениус, Генрик Антон Лоренц) сочли нужным написать элементарные пособия по высшей математике «для естествоиспытателей и врачей», как сказано было в заголовке книги Аррениуса, или «для естествоиспытателей с особым вниманием к запросам химиков», как указывал читателям Нернст. Из этих книг в нашей стране лучше всего известен двухтомный «Курс дифференциального и интегрального исчисления с их приложениями к естествознанию» Г. А. Лоренца [4] (или

«Курс высшей математики», как назвал русский вариант книги ее переводчик В. П. Шереметьевский), выдержавший с 1901 по 1926 г. пять изданий и еще в 30-х годах весьма популярный как учебник для будущих инженеров и пособие для самообразования. При этом с идейной стороны все названные книги были достаточно близки к настоящей книге.

Следует указать, что и многие «чистые» математики, не специализирующиеся по абстрактной алгебре или математической логике и имеющие вкус к приложениям, мыслят аналогично тем физикам, о которых упоминалось выше. Так, знаменитый немецкий (позже американский) математик Рихард Курант (1888—1972), основатель Курантовского математического института в Нью-Йорке, писал в 1964 г., что очень долго математики принимали геометрию Евклида за образец строго логического подхода, строго логической дедукции (вывода). Но вот что говорит Курант дальше [5]:

> «Упор на этот [аксиоматический, логический] аспект полностью дезориентирует того, кто предположит, что созидание, воображение, сопоставление и интуиция играют только вспомогательную роль в математическом творчестве и в настоящем понимании. В математическом образовании дедуктивный метод, начинающийся с догматически сообщаемых аксиом, действительно позволяет быстрее обозреть большую территорию. Но конструктивный метод, идущий от частного к общему и избегающий догматического принуждения, надежнее ведет к самостоятельному творческому мышлению».

Все сказанное выше достаточно полно поясняет основные установки нашей книги. Ее принципы и исходные положения совпадают с теми, которые были характерны для названного в сноске 1 учебника; в деталях же изложение существенно отличается от старого даже и в тех разделах, которые входили в названную выше книгу. Многие темы введены в нашу книгу заново; это касается как ряда частных моментов (вопрос о втором дифференциале функции одной переменной или о полном дифференциале функций двух переменных; тема об инерционных системах отсчета), так и целых больших направлений науки (ряды Фурье, теория функций комплексной переменной или устройство лазеров).

При пользовании этой книгой в преподавании педагог может свободно использовать те или иные моменты из физических разделов книги для иллюстраций математических понятий и теорем или, напротив, включать отдельные математические темы в занятия по физике. Так, например, понятие производной мы иллюстрировали в тексте на связи пройденного пути и скорости или требующегося для нагревания тела тепла и теплоемкости; но никто не мешает использовать здесь, скажем, связь между электродвижущей силой (ЭДС) и силой тока для катушки (или для конденсатора, где эта связь имеет обратный характер; ср. § 13.1). Разумеется, число примеров подобного рода можно неограниченно увеличить. Слитное преподавание математического анализа и его приложений будущим техникам и физикам мы считаем очень важным: ведь сама математическая техника наполняется глубоким содержанием лишь в процессе ее использования. Польза от физических разделов книги является двоякой: они не только доставляют учащимся важные сведения по физике, но также и проливают дополнительный свет на математические построения первых глав, позволяют по-новому понять их естественность и необходимость. При преподавании часть этого материала можно, конечно, и опустить; однако какое-то использование физических рассмотрений в упражнениях (а частично, и в лекциях) по математике кажется нам очень полезным, а серьезное согласование элементарных курсов математики и физики — совершенно необходимым.

Я. Б. Зельдович,
И. М. Яглом

---

[5] С тех же позиций написана и превосходная книга Дж. Пойа «Математика и правдоподобные рассуждения» (М.: Мир, 1977).

ЧАСТЬ ПЕРВАЯ
ЭЛЕМЕНТЫ ВЫСШЕЙ МАТЕМАТИКИ

Глава 1
ФУНКЦИИ И ГРАФИКИ

Настоящая глава книги является вводной — в ней кратко повторяются некоторые сведения, вероятно известные большинству читателей. Эта глава посвящена функциональным зависимостям между величинами; в ней приводятся некоторые конкретные примеры функциональных зависимостей, часто используемые в последующих разделах книги. С другими важными примерами функциональных зависимостей читатель встретится в последующих главах.

§1. Функциональная зависимость

В природе и в технике мы очень часто встречаемся с зависимостями одних величин от других — с так называемыми *функциональными зависимостями*; естественно, что и математика уделяет таким зависимостям большое внимание. Функциональная зависимость одной величины (y) от другой (x) означает, что каждому значению x соответствует определенное значение y. Величина x при этом называется *независимой переменной*, а y — *функцией* этой переменной. Иногда x называют *аргументом* функции y.

Приведем несколько примеров из геометрии и физики.

1. Площадь S квадрата является функцией длины a его стороны: $S=a^2$.

2. Объем V шара можно выразить через радиус R шара: $V=\frac{4}{3}\pi R^3$.

3. Объем V конуса с данной высотой h зависит от радиуса r его основания:

$$V=\frac{1}{3}\pi r^2 h; \tag{1.1.1}$$

если же считать, напротив, известным r, то та же формула (1) выразит зависимость объема V конуса от его высоты h.

4. Путь z, пройденный свободно падающим телом, зависит от времени t, протекшего с момента, когда началось падение. Эта зависимость выражается формулой

$$z=\frac{gt^2}{2}, \tag{1.1.2}$$

где $g\approx 9{,}8$ м/с² — ускорение свободного падения.

5. Сила тока i зависит от сопротивления R проводника: при данной разности потенциалов u

$$i=\frac{u}{R}. \tag{1.1.3}$$

Можно было бы привести еще множество примеров такого рода.

В математике функциональная зависимость чаще всего задается формулами, например:

$$y=2x+3,\quad y=x^2+5,$$
$$y=3x^3-x^2-x, \tag{1.1.4}$$
$$y=\frac{x-1}{x+1}\ \text{ или }\ y=\sqrt{3x+7}$$

(здесь всюду y — функция аргумента x). Формула дает способ вычисления значения функции при каждом заданном значении независимой переменной. Иногда этот способ предполагает использование таблиц или иных технических средств. Так, в последнем из примеров (4) для вычисления значения функции нам придется извлечь квадратный корень из числа, скажем, с помощью таблицы квадратных корней или с помощью специального вычислительного устройства.

В физике и технике функциональные зависимости чаще всего фиксируются на шкалах измерительных приборов; так, движение стрелки спидометра автомобиля, наблюдаемое одновременно со стрелками Ваших часов, фиксирует функциональную зависимость скорости автомобиля от времени. Но и здесь часто функцию можно с достаточной степенью точности записать простой

формулой: например, если разность потенциалов в электрической цепи остается постоянной, то наблюдаемая экспериментально («считываемая» со шкалы амперметра) зависимость между силой тока i и сопротивлением цепи R хорошо передается формулой (3). Наличие формулы, связывающей значения возникающих в физике или технике величин, означает, что получен закон интересующего нас явления; так, выражаемый формулой (3) физический закон называется *законом Ома*.

Реально в природе и в технике в большинстве случаев интересующая нас величина (функция) зависит от нескольких величин. В последнем примере (см. (3)) сила тока зависит от двух величин: от разности потенциалов u и от сопротивления проводника R. Объем конуса (см. (1)) зависит от его высоты h и от радиуса основания r. Считая заданными (и постоянными) все величины, кроме одной, мы изучаем зависимость функции от одной переменной; в данной книге мы в основном ограничимся функциями одной переменной. Так, например, взяв данную аккумуляторную батарею с определенной разностью потенциалов u, будем менять сопротивление проводника R и измерять силу тока i. В такой постановке опыта сила тока зависит только от сопротивления; величину u в формуле (3) следует рассматривать как постоянный коэффициент.

Иногда формула такова, что значение y существует только для x, заключенного в определенных пределах. Скажем, формула $y=\sqrt{x}$ позволяет найти y только для неотрицательных значений x, т. е. для $x \geqslant 0$. Если $y=\log_2(x-2)$, то y существует только для $x-2>0$, т. е. для $x>2$. В предпоследней из формул (4) приходится считать, что $x\neq-1$, а в последней — что $3x+7\geqslant 0$, т. е. $x\geqslant -{}^7/_3\approx-2{,}33$.

В формулах, возникающих в физических, технических, геометрических и других задачах, ограничения возможных значений независимой переменной иногда вытекают из смысла рассматриваемой задачи. В примерах, приведенных в начале параграфа, при вычислении площади квадрата и объема конуса, конечно, подразумевается, что сторона квадрата, а также радиус основания конуса и высота конуса являются положительными величинами. При этом «умные» формулы зачастую сами предупреждают нас о возможных границах значений переменной. Так, формула для объема конуса при отрицательных h дает отрицательное значение объема, что противоестественно; во многих формулах теории относительности скорость v движущегося тела фигурирует в виде выражения $\sqrt{c^2-v^2}$, где c — скорость света, откуда вытекает, что в теории относительности всегда $v<c$, и т. д.

В первых трех формулах (4) и, конечно, во многих других никаких ограничений на величину независимой переменной не накладывается: x может быть положительным или отрицательным, большим или малым. То же относится и ко многим физическим величинам: например, электрический ток удобно считать положительной величиной, когда электроны преимущественно движутся по проволоке в одном направлении, и отрицательной при противоположном направлении движения электронов.

Зная формулу, дающую зависимость y от x, легко составить таблицу значений y для нескольких произвольно заданных значений x.

Составим таблицу, отвечающую функции $y=3x^3-x^2-x$. В верхней строке приведем выбранные нами значения x, в нижней строке под каждым данным x выпишем соответствующее значение y:

x	-3	-2	-1	0	1	2	3
$y=3x^3-x^2-x$	-87	-26	-3	0	1	18	69

Понятно, что по данной формуле можно составить и более подробную таблицу, задавая, например, значения $x=0$; 0,1; 0,2; . . . Таким образом, формула «сильнее» любой таблицы. Формула содержит не только те сведения, которые приведены в данной таблице, но позволяет найти значения функции также и при значениях независимой переменной, не содержащихся в таблице. С другой стороны, таблица удобнее формулы, так как с ее помощью можно быстрее найти значение y при данном x, если это значение x есть в таблице. Таблица также «нагляднее»

сложной формулы, по которой зачастую трудно оценить значения, принимаемые функцией; однако простая формула, при некоторой привычке в чтении таких формул, часто позволяет быстрее представить себе ход функции, чем невыразительный ряд чисел (таблица).

В естественных науках и в технике часто встречается такое положение, когда теории интересующего нас явления еще нет и физик (или химик, биолог, техник) может указать только результаты проделанных им опытов — зависимость исследуемой величины от величины, задаваемой при постановке опыта. Так обстоит дело, например, при исследовании зависимости сопротивления проводника от его температуры. В этом случае функциональная зависимость может быть задана только в виде таблицы, суммирующей результаты опыта.

Из опыта известно, что для данного проводника (из данного материала, данного сечения и данной длины) электрическое сопротивление зависит от температуры проводника. При каждом значении температуры T проводник имеет определенное сопротивление R, так что можно говорить о том, что сопротивление R есть функция температуры T. Проводя измерения, можно найти значения R при различных T и таким образом найти зависимость $R(T)$; при этом результатом опытов является таблица, в которой даны значения R при различных T, например, следующая:

T (в °C)	0	25	50	75	100
R (в Ом)	112,0	118,4	124,6	130,3	135,2

Если нас интересуют значения R при температурах, не входящих в таблицу, то в принципе нужны дополнительные измерения, так как точная формула, дающая зависимость R от T, неизвестна. Однако практически всегда можно подобрать приближенную формулу, хорошо согласующуюся с опытом для тех температур, при которых проведены измерения. Возьмем, например, соотношение

$$R = 112{,}0 + 0{,}272T - 0{,}0004T^2 \qquad (1.1.5)$$

и составим по этой формуле следующую таблицу:

T (в °C)	0	25	50	75	100
R (в Ом)	112,0	118,55	124,6	130,15	135,2

При тех температурах, при которых проделаны измерения, формула (5) дает значения R, очень близкие к наблюдаемым на опыте. Поэтому законно предположение, что и при промежуточных температурах (например, при $T = 10$, 80 или 90° C) эта формула также правильно описывает функциональную зависимость $R(T)$. Математики формулируют последнее утверждение, говоря, что найденная зависимость $R(T)$ не дает больших ошибок при *интерполяции*, т. е. при переходе от известных значений R к новым, промежуточным между уже имеющимися (латинское слово interpolare означает «подновлять»). При этом зависимость (5) сопротивления R от температуры T называют эмпирически найденным законом, или *эмпирической формулой*. (Прилагательное «эмпирический» (от греческого слова empeiria — опыт) означает «опытный», «полученный опытным путем».) Однако эмпирическая формула нуждается, разумеется, в проверке: погрешность, получаемая при ее использовании, может оказаться и довольно значительной (ясно, что надежность эмпирической формулы будет тем выше, чем более густой является сетка тех наблюдаемых значений переменной, исходя из которых мы подбирали данную формулу). И совсем уж нежелательно использование эмпирической формулы за пределами исследованного интервала значений независимой переменной (такое продолжение формулы называется *экстраполяцией* — от латинской частицы ex, т. е. «вне»): это может привести к большим ошибкам. Так, например, мы не можем применять формулу (5) при $T = -200°$ C или при $T = 500°$ C, поскольку у нас нет никаких оснований ожидать, что при в с е х T зависимость $R(T)$ имеет вид (5), т. е. является *квадратичной* (выражается квадратным трехчленом).

§ 2. Координаты. Расстояния и углы, выраженные в координатах

Для наглядного изображения функциональной зависимости с помощью рисунка (графика) пользуются (прямоугольными декартовыми) *координатами*. Проведем на плоскости две взаимно перпендикулярные прямые: ось x, или *ось абсцисс*, которая обычно располагается горизонтально, и ось y, или *ось ординат*, которую проводят вертикально; точка O пересечения осей называется *началом координат*. Обычно представляют себе, что плоскость, на которой проведены оси x и y, поставлена перед Вами вертикально, как стена, против которой Вы сидите. При этом ось x оканчивается стрелкой, направленной слева направо, а ось y — стрелкой, направленной снизу вверх. Эти стрелки указывают, что отрезок OP оси x считается положительным, если он идет от O к P слева направо, и отрицательным, когда P лежит левее O; аналогично, отрезок OQ оси y будет положителен, если Q лежит выше O, и отрицателен в обратном случае.

Каждой паре значений x и y, скажем $x=2$, $y=4$, отвечает на координатной плоскости одна точка. Для того чтобы ее построить, достаточно отложить по осям абсцисс и ординат отрезки $OB=2$

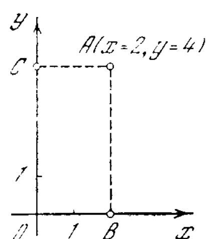


Рис. 1.2.1

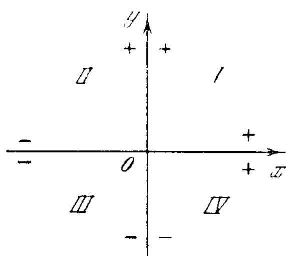


Рис. 1.2.2

и $OC=4$ (причем эти отрезки направлены вправо и вверх соответственно, ибо у нас и x и y положительны (рис. 1)). Восставив в точках B и C перпендикуляры к осям координат Ox и Oy, мы получим в пересечении этих осей точку A ($x=2$, $y=4$), или, короче, точку A (2, 4) (рис. 1). Обратно, чтобы найти координаты произвольной точки A плоскости, достаточно мысленно опустить из A перпендикуляры AB и AC на оси координат и «прочесть» значения x и y, отвечающие точкам B и C (эти значения равны длинам отрезков OB и OC, взятым с соответствующими знаками). При этом в четырех четвертях (или «квадрантах») I, II, III, IV,

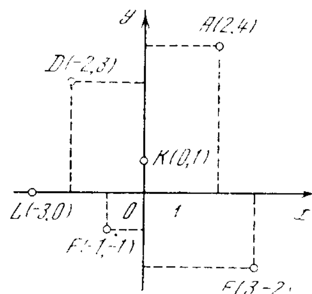


Рис. 1.2.3

на которые оси координат делят плоскость (рис. 2), знаки координат будут таковы: (+, +) (т. е. положительная абсцисса x и положительная ордината y), (—, +), (—, —) и (+, —) соответственно. Оси абсцисс Ox отвечает значение $y=0$, а оси ординат — значение $x=0$; начало координат O имеет координаты (0, 0). На рис. 3 изображены несколько точек и выписаны их координаты.

Полезно научиться быстро определять на глаз приближенные значения координат заданных на рисунках точек и легко находить точки, отвечающие тем или иным известным нам значениям координат. Практически удобно находить точки и строить графики с помощью миллиметровой бумаги, на которой уже заранее проведена сетка взаимно перпендикулярных линий. Так, например, мы рекомендуем читателю постараться, не прибегая ни к каким вспомогательным построениям, назвать координаты изображенных на рис. 4 точек A—N.

Задание двух чисел — значений x и y — определяет положение точки на плоскости. Поэтому и все геометрические величины, относящиеся к этой точке, можно выразить через координаты точки.

Найдем, например, *расстояние* r точки A с координатами x и y от начала координат, т. е. длину r отрезка OA, соединяющего начало координат O с точкой A (рис. 5), а также угол α между прямой OA и осью абсцисс.

Опустим из A перпендикуляры AB и AC на оси координат. Длина отрезка OB равна $|x|$, длина отрезка AB равна

длине OC, т. е. $|y|$. Из прямоугольного треугольника OAB по теореме Пифагора находим

$$(OA)^2 = r^2 = (OB)^2 + (AB)^2 = x^2 + y^2.$$

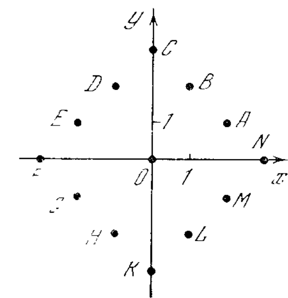


Рис. 1.2.4

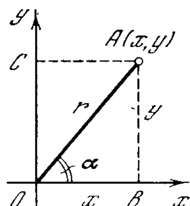


Рис. 1.2.5

Таким образом,

$$r = \sqrt{x^2 + y^2}. \tag{1.2.1}$$

Далее, по определению тангенса угла, имеем

$$\operatorname{tg} \alpha = \frac{AB}{OB} = \frac{y}{x}. \tag{1.2.2}$$

Так, например, пусть $x=2$, $y=3$ (см. рис. 5). Тогда

$$r = \sqrt{13} \approx 3{,}6, \quad \alpha = \operatorname{arc\,tg} \frac{3}{2} \approx 56°.$$

Заметим, что угол α отсчитывается от **положительного** направления оси x в направлении, **противоположном** движению часовой

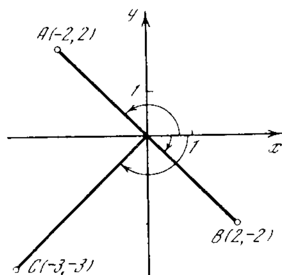


Рис. 1.2.6

стрелки. Поэтому, например, если $y=2$, $x=-2$ (см. точку A на рис. 6), то угол α — тупой, $\operatorname{tg} \alpha = \frac{2}{-2} = -1$, $\alpha = 135°$. Когда точка лежит ниже оси x, то угол α оказывается большим 180°: так, для точек B (2, —2) и C (—3, —3) (см. рис. 6) имеем $\alpha=315°$ (точка B) и $\alpha=225°$ (точка C). Так как столь большие углы иногда оказываются неудобными, то в этих случаях можно отсчитывать углы от положительного направления оси x в направлении, **совпадающем** с направлением движения часовой стрелки, но считая их при этом **отрицательными**; при таком соглашении для точки B имеем $\alpha=-45°$, а для точки C получаем $\alpha=-135°$. Другими словами, мы можем считать, что угол α изменяется либо в пределах от 0 до 360° (причем значения 360° он не принимает, ибо тут удобнее считать $\alpha=0°$), либо от —180 до 180°, причем значения $\alpha = -180°$ и $\alpha=180°$ отвечают точкам, принадлежащим «отрицательной полуоси» оси x (причем здесь законны оба эти значения α).

Формула (2) в этом отношении не полна: она не позволяет установить, расположена ли рассматриваемая точка в I или в III четверти координатной плоскости (или попадает ли она во II или в IV четверть), поскольку в этих четвертях тангенс имеет один и тот же знак. Для того чтобы полностью определить величину α, надо еще учесть знаки x и y, позволяющие определить четверть, в которой находится точка, или воспользоваться более полными, чем (2), но и более сложными формулами:

$$\cos \alpha = \frac{x}{r}, \quad \sin \alpha = \frac{y}{r}, \tag{1.2.2а}$$

где $r = \sqrt{x^2 + y^2}$.

Легко решить и обратную задачу: пусть дано, что точка A находится на заданном расстоянии r от начала координат O и отрезок OA образует угол α с осью x (подразумевается — с положительной, правой частью оси x). Требуется найти координаты точки A. Из рис. 5 получаем:

$$x = r \cos \alpha, \quad y = r \sin \alpha. \tag{1.2.3}$$

Эти формулы верны без всяких исключений, для любых положительных и отрицательных углов α: они правильно указывают знаки x и y в любой четверти.

Ясно, что положение точки A на плоскости можно фиксировать, указав два числа x и y, а можно вместо этого задать расстояние r и угол α. Числа r и α называют *полярными координатами* точки A; при этом точку O называют *полюсом* полярной системы координат, а луч Ox — *полярной осью* (ось Oy

в определении полярной системы координат не участвует). Таким образом, формулы (2а) и (3) указывают правила перехода от декартовых прямоугольных координат точки к соответствующим им полярным и наоборот.

Перейдем к задачам, связанным с двумя точками A_1 и A_2; координаты первой точки обозначим x_1, y_1, координаты второй точки — x_2, y_2 (рис. 7). Найдем *расстояние* r_{12} между этими точками и угол α_{12} между отрезком A_1A_2 и осью x [1].

Удобно провести через точку A_1 прямую, параллельную оси x, а через точку A_2 — прямую, параллельную оси y (на рис. 7 они проведены пунктиром, а точка их пересечения обозначена буквой B). При этом в треугольнике A_1A_2B отрезок A_1B равен $|x_2-x_1|$, а отрезок A_2B равен $|y_2-y_1|$. Построение ΔA_1A_2B подобно построению ΔOAB на рис. 5.

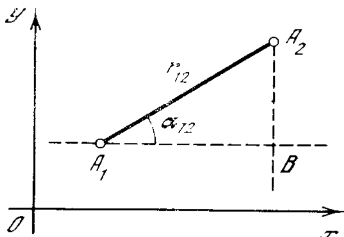


Рис. 1.2.7

По теореме Пифагора

$$r_{12}=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}. \qquad (1.2.4)$$

Угол α_{12} находится из условия

$$\operatorname{tg}\alpha_{12}=\frac{y_2-y_1}{x_2-x_1}, \qquad (1.2.5)$$

или (ср. (2а)) из

$$\cos\alpha_{12}=\frac{x_2-x_1}{r_{12}}=\frac{x_2-x_1}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}, \qquad (1.2.5а)$$

$$\sin\alpha_{12}=\frac{y_2-y_1}{r_{12}}=\frac{y_2-y_1}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}.$$

Читатель должен убедиться, что формулы (4), (5) и (5а) правильны при любых знаках всех четырех величин x_1, y_1, x_2, y_2 и при любых соотношениях между координатами: может быть

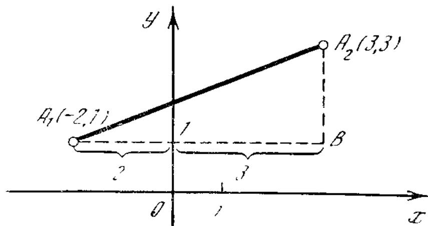


Рис. 1.2.8

$x_1>x_2$ или $x_1<x_2$, $y_1>y_2$ или $y_1<y_2$. Так, например, рис. 8 иллюстрирует случай $x_1<0$, $x_2>0$. Здесь $A_1=A_1(-2, 1)$ и $A_2=A_2(3, 3)$. В этом случае длина отрезка A_1B равна сумме абсолютных величин $|x_1|=2$, $|x_2|=3$, но это как раз и соответствует общей формуле

$$A_1B=x_2-x_1=3-(-2)=5.$$

Аналогично правильны и выражения для r_{12} и $\operatorname{tg}\alpha$.

Упражнения

1.2.1. Изобразите точки (1, 1); (—1, 1); (—1, —1); (1, —1).

1.2.2. Изобразите точки (1, 5); (5, 1); (—1, 5); (—5, 1); (—1, —5); (—5, —1); (1, —5); (5, —1).

1.2.3. Изобразите точки (0, 4); (0, —4); (4, 0); (—4, 0).

1.2.4. Найдите расстояние от начала координат и угол α для точек (1, 1); (2, —2); (—3, —3); (—4, 4).

1.2.5. Найдите расстояние между парами точек: A_1 (1, 1) и A_2 (1, —1); A_1 (1, 1) и A_2 (—1, —1); A_1 (2, 4) и A_2 (4, 2); A_1 (—2, —4) и A_2 (—4, —2).

1.2.6. Выпишите координаты вершин квадрата со стороной a, если диагонали квадрата совпадают с осями x и y.

1.2.7. Выпишите координаты вершин правильного шестиугольника со стороной a, если одна из его диагоналей совпадает с осью x, а центр лежит в начале координат.

1.2.8. а) Выпишите координаты вершин равностороннего треугольника со стороной a,

[1] Значки снизу букв называются *индексами* (по-латыни «индекс» значит «указатель»). Читается такое выражение: x_1 — икс-один, A_2 — а-два. Одинаковые буквы с разными индексами применяют вместо разных букв для того, чтобы показать, что речь идет о похожих (но в то же время разных) величинах. Так, например, величины x_1 и x_2 обе являются «иксами» — абсциссами, но относятся к разным точкам. Величины, обозначаемые разными буквами, но с одинаковым индексом, относятся к одной и той же точке: A_1 — обозначение некоторой точки, x_1 — ее абсцисса, y_1 — ордината. Иногда при букве ставят и два индекса: так, например, r_{12} (читается: эр-один-два, а не эр-двенадцать) — это расстояние между первой (A_1) и второй (A_2) точками.

основанием, принадлежащим оси x, и вершиной противолежащего угла на оси y. б) Решите ту же задачу для случая, когда основание принадлежит оси x, а вершина одного из углов совпадает с началом координат.

1.2.9. Дана точка A_1 с координатами x_1, y_1. Выпишите координаты точек A_2, A_3 и A_4, симметричных A_1 относительно соответственно осей x и y и относительно начала координат O.

§ 3. Графическое изображение функций. Уравнение прямой

Как мы знаем из § 2, каждой паре значений величин x и y соответствует определенная точка плоскости. Если y есть функция от x, то каждому значению x отвечает определенное значение y. Следовательно, задаваясь многими различными значениями x, мы найдем различные соответствующие им y, и эти пары значений (x, y) дадут много точек на плоскости. Если увеличивать число отдельных значений x, беря их все более близкими между собой, то в конце концов точки сливаются в сплошную кривую. Эта кривая называется *графиком* функции.

Обычно достаточно нанести на график несколько точек, после чего промежуточные точки и весь график функции можно получить, соединяя плавной линией нанесенные точки. Однако

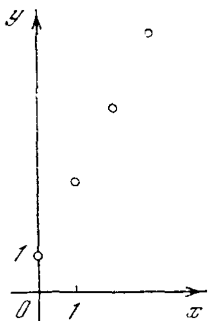


Рис. 1.3.1

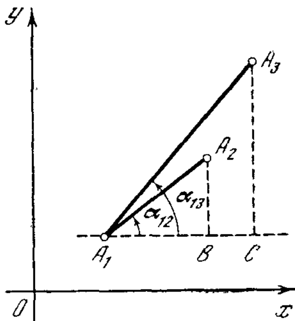


Рис. 1.3.2

для того чтобы не сделать при этом грубых ошибок, нужно иметь общее представление о виде кривых, изображающих различные функции. Мы начнем с изучения нескольких типичных и наиболее важных функций.

Рассмотрим так называемую *линейную* зависимость

$$y = kx + b. \qquad (1.3.1)$$

Пусть, например,

$$y = 2x + 1.$$

Построим несколько точек графика линейной зависимости; соответствующие значения x и y приведены ниже:

x	0	1	2	3
y	1	3	5	7

Изобразим эти точки на плоскости (рис. 1). Бросается в глаза, что все наши точки лежат на одной прямой. В таком случае, проводя эту прямую линию (отсюда название «линейная зависимость», «линейная функция»), получим полный график функции: для любого x соответствующая точка (x, y) принадлежит прямой, соединяющей какие-либо две точки графика.

Как доказать, что для любой функции вида $y=kx+b$ (при любых k и b) все точки ее графика принадлежат одной прямой? Поставим следующий вопрос: как, зная координаты трех точек A_1, A_2 и A_3, выяснить (без построения, путем вычисления), лежат ли эти точки на одной прямой или нет? Очевидно, что если угол α_{12} отрезка A_1A_2 с осью x равен углу α_{13} отрезка A_1A_3 с осью x, то отрезки A_1A_2 и A_1A_3 принадлежат одной прямой, а если это не так, то не принадлежат. На рис. 2 показан случай, когда $\alpha_{13} > \alpha_{12}$ и точка A_3 лежит выше продолжения прямой A_1A_2; из этого же рисунка видно, что если бы α_{13} было равно α_{12}, то точка A_3 лежала бы на прямой, продолжающей отрезок A_1A_2.

Из выражения (2.5) для тангенса угла α следует, что при $\alpha_{12}=\alpha_{13}$ координаты точек A_1, A_2 и A_3 связаны соотношением

$$\frac{y_2-y_1}{x_2-x_1}=\frac{y_3-y_1}{x_3-x_1}. \qquad (1.3.2)$$

Не обращаясь к тригонометрии, можно сказать, что условие (2) — это условие подобия двух прямоугольных треугольников A_1A_2B и A_1A_3C (см. рис. 2). Подобие треугольников и приводит к равенству углов при вершине A_1.

Соотношение (2) применимо и в том случае, когда точка A_1 лежит между точками A_2 и A_3 (рис. 3): если три точки лежат на одной прямой, то из

подобия ΔA_1A_2B и ΔA_1A_3C следует пропорция (2). В примере, изображенном на рис. 3, $x_3-x_1<0$, $y_3-y_1<0$, но отношение их положительно и равно отношению двух положительных величин x_2-x_1 и y_2-y_1.

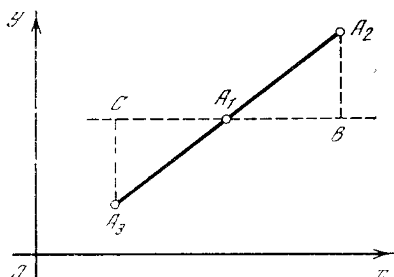


Рис. 1.3.3

Теперь проверим, что условие (2) выполняется для любой тройки точек графика линейной функции (1). Рассмотрим две точки $A(x_1, y_1)$ и $B(x_2, y_2)$, координаты которых удовлетворяют уравнению (1). В этом случае $y_1=kx_1+b$, а $y_2=kx_2+b$ и $y_2-y_1=kx_2+b-(kx_1+b)=k(x_2-x_1)$, откуда

$$\frac{y_2-y_1}{x_2-x_1}=k.$$

Отношение оказалось не зависящим от x_1 и x_2. Следовательно, и для любой другой пары точек графика, в частности для пары точек $A(x_1, y_1)$ и $C(x_3, y_3)$, получим также

$$\frac{y_3-y_1}{x_3-x_1}=k=\frac{y_2-y_1}{x_2-x_1}.$$

Значит, для любых трех точек графика $A(x_1, y_1)$, $B(x_2, y_2)$ и $C(x_3, y_3)$ имеет место соотношение (2), т. е. любые три точки лежат на одной прямой, а, следовательно, одной прямой принадлежат в с е точки графика линейной функции $y=kx+b$. Значит, графиком функции $y=kx+b$ является прямая линия; эту линию мы часто будем кратко называть «прямой $y=kx+b$» (в нашем случае ее также можно назвать «линией (или прямой) (1)»).

Уравнение $y=kx+b$ называют *уравнением прямой*. Коэффициент k определяет угол между прямой и осью x. Подставляя в уравнение $x=0$, получим $y=b$. Значит, одна из точек прямой — это точка $(0, b)$; эта точка лежит на оси y на высоте b над началом координат (если $b<0$, то точка лежит под началом координат). Таким образом, b — это ордината точки пересечения прямой с осью y (иногда величину b называют «начальной ординатой» прямой), $|b|$ — длина отрезка, отсекаемого прямой на оси ординат (в случае, изображенном на рис. 1, $b=1$).

Для того чтобы построить прямую, соответствующую данному уравнению, не нужно вычислять координаты большого числа точек и наносить их на график: ясно, что если построены две точки, то тем самым полностью определена прямая, проходящая через эти точки. Можно, например, всегда брать две точки ($x=0$, $y=b$) и ($x=1$, $y=b+k$) и проводить прямую по этим точкам. Можно в качестве второй точки взять точку пересечения прямой с осью x (т. е. точку $x=x_0$, $y=0$). Из условия $y=kx_0+b=0$ найдем $x_0=-\frac{b}{k}$.

Полезно так поупражняться в построении графиков, чтобы сразу, с одного взгляда на уравнение, примерно представить себе общий вид соответствующей линии, так сказать, «ход» интересующей нас функции.

Для линейной функции, графиком которой является прямая линия, это совсем легко. Ведь на самом деле ход прямой зависит только от двух величин k и b, входящих в уравнение прямой. Таким образом, нужно разобрать не так уж много вариантов: k может быть положительно или отрицательно; k может быть большим или малым по абсолютной величине (больше 1 или меньше 1); b может быть положительным или отрицательным.

Покажем, как проводится такое исследование.

Начнем со случая $b=0$, т. е. с уравнения $y=kx$. Соответствующая прямая, очевидно, проходит через начало координат, т. е. через точку $x=0$, $y=0$. На рис. 4 показано несколько прямых с различными значениями k, обозначенными на обоих концах каждой прямой. Проверьте правильность проведения каждой прямой, и тогда Вы убедитесь в правильности следующих общих выводов:

1) если $k>0$, то прямая лежит в I и III четвертях; если $k<0$ — во II и IV четвертях;

2) при $k=1$ прямая лежит в I и III четвертях, причем луч, принадлежащий I четверти, образует угол $\alpha=45°$ с осью x, т. е. делит пополам угол между осью x и осью y (напомним, что под словами «угол с осью x» подразумевается угол с положительным направлением оси x, показанным стрелкой). Продолжение прямой, лежащее

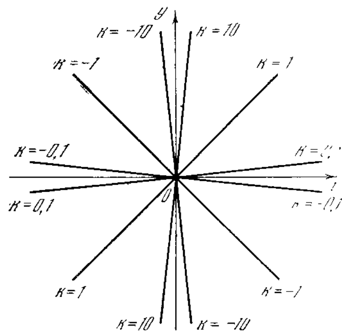


Рис. 1.3.4

в III четверти, образует с осью x угол $\alpha=-135°$ — в этом случае прямая является *биссектрисой* I *и* III *координатных углов*;

3) при $k=-1$ часть прямой, лежащая во II четверти, образует угол $\alpha=135°$ с осью x, а продолжение прямой в IV четверти — угол $\alpha=-45°$; прямая является *биссектрисой* II и IV *координатных углов*;

4) если $|k|<1$, прямая идет полого, т. е. ближе к оси x, чем к оси y, и тем более полого, чем меньше $|k|$. Если $|k|>1$, прямая идет круто: ближе к оси y, чем к оси x (тем круче, чем больше $|k|$).

Число k, характеризующее «крутизну» прямой, задаваемую углом α, образованным прямой с осью x, называется *угловым коэффициентом*, или *наклоном*, прямой; мы чаще будем употреблять краткий термин «наклон».

Теперь, когда это усвоено, мы можем перейти к общему случаю: $b\neq 0$. Пусть имеются две прямые: прямая $y=kx$ (например, $y=0,5x$; рис. 5), в уравнении (1) которой $b=0$, и прямая с тем же наклоном k, но с $b\neq 0$ (например, прямая $y=0,5x+2$). Для удобства мы будем писать [1]: $y_0=0,5x$, $y_2=0,5x+2$. При каждом данном x величина y_2 на две единицы больше y_0. Значит, точки прямой y_2 получаются из точек прямой y_0 с тем же x сдвигом на две единицы вверх. Поэтому прямая y_2 параллельна прямой y_0 и лежит в ы ш е нее на две единицы. Очевидно, что такое правило справедливо при любом b (если $b<0$, то прямая лежит

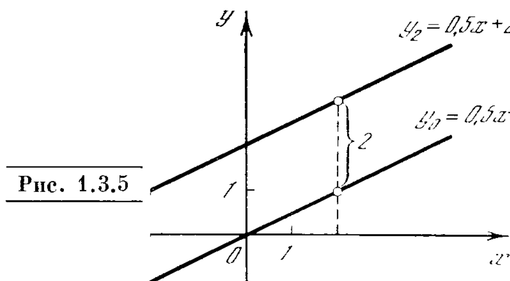


Рис. 1.3.5

н и ж е начала координат, ниже соответствующей прямой $y=kx$).

Усвоив, как идут прямые с уравнениями $y=kx$ при различных k, мы теперь легко представим себе общий ход прямой $y=kx+b$ с любыми k и b. Упражнения для тренировки даны в конце параграфа. В частном случае $k=0$ получается уравнение $y=b$ (подразумевается, что $y=b$ при л ю б ы х значениях x), чему соответствует горизонтальная прямая, наклон которой равен нулю (рис. 6).

Можно представить себе пешехода, идущего по прямой слева направо,

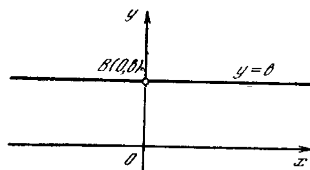


Рис. 1.3.6

в сторону возрастающих значений x. Если $k>0$, то пешеход поднимается в гору (наклон положителен); если $k<0$, пешеход идет под гору (отрицательный наклон); если $k=0$ (случай нулевого наклона), то путь пешехода проходит по ровной местности — он горизонтален.

Наклон k указывает отношение изменения функции к соответствующему изменению аргумента функции. В самом деле, для любых двух точек

[1] Здесь мы пользуемся индексами несколько иначе, чем раньше. Величина y_0 относится к целой прямой, это не ордината определенной точки, а ордината произвольной точки определенной прямой — той, которой отвечает значение $b=0$; точно так же запись y_2 относится к прямой с начальной ординатой $b=2$. Таким образом, здесь y_0 и y_2 — это не числа, а функции; $y_0=y_0(x)$ и $y_2=y_2(x)$.

$A\ (x_1, y_1)$ и $B\ (x_2, y_2)$ графика линейной функции имеем

$$\frac{y(x_2)-y(x_1)}{x_2-x_1}=\frac{kx_2+b-(kx_1+b)}{x_2-x_1}=k.$$

Такое отношение мы уже вычисляли выше, когда доказывали, что линейная функция на графике изображается прямой линией. В общем случае произволь-

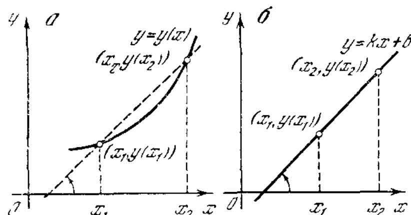


Рис. 1.3.7

ной функции $y(x)$ аналогичная величина $\frac{y(x_2)-y(x_1)}{x_2-x_1}$ равна тангенсу угла между отрезком, соединяющим точки $(x_1, y(x_1))$ и $(x_2, y(x_2))$, и осью абсцисс (рис. 7, *а*).

Линейная функция отличается тем, что для нее эта величина одинакова для любой пары точек: она не зависит ни от x_2, ни от x_1 (рис. 7, *б*). Поэтому и получается, что все точки графика ли-

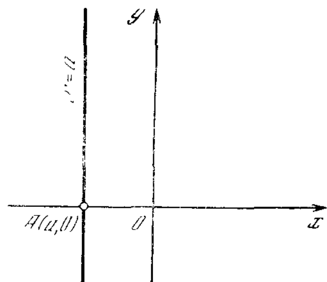


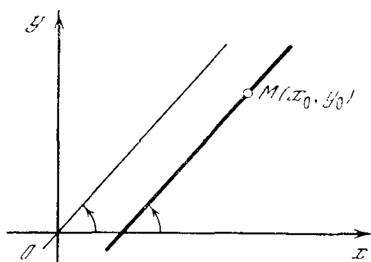
Рис. 1.3.8

нейной функции принадлежат одной прямой.

Заметим еще, что прямая, параллельная оси y (такая прямая не является графиком никакой функции), записывается уравнением $x=a$ (подразумевается, что $x=a$, а y может быть каким угодно; см. рис. 8, где $a<0$). Угловой коэффициент k такой прямой уместно считать *бесконечным* ($k=\infty$): ведь в этом случае отношение $k=(y_2-y_1)/(x_2-x_1)$ где (x_1,y_1) и (x_2,y_2) — точки нашей прямой, будет иметь вид $c/0$ (прямая $x=a$ — прямая бесконечного наклона).

Теперь легко найти уравнение прямой, *проходящей через данную точку* $M(x_0, y_0)$ *и имеющей известный наклон* k. Такое уравнение будет иметь вид (1) (где k нам известно, а b — пока нет), причем ему должны удовлетворять значения $x=x_0$, $y=y_0$ (рис. 9). Это уже

Рис. 1.3.9



подсказывает, как будет выглядеть требуемое уравнение — его вид таков:

$$y-y_0=k(x-x_0), \qquad (1.3.3)$$

или

$$y=kx+(y_0-kx_0). \qquad (1.3.3а)$$

В самом деле, наклон прямой (3а), или (3), равен k; с другой стороны, подстановка в обе части (3) значений $x=x_0$ и $y=y_0$ обращает это равенство в тождество $0=0$.

Упражнения

1.3.1. Выясните, лежат ли на одной прямой или нет тройки точек: $A_1\ (0,\ 0)$, $A_2\ (2,\ 3)$, $A_3\ (4,\ 6)$; $A_1\ (0,\ 0)$, $A_2\ (2,\ 3)$, $A_3\ (-2,\ -3)$; $A_1\ (2,\ -3)$, $A_2\ (4,\ -6)$, $A_3\ (-2,\ 3)$.

1.3.2. Постройте прямые: $y=3x$; $y=3x+2$; $y=3x-1$; $y=2-x$; $y=2-0{,}5x$; $y=x-3$.

1.3.3. Найдите уравнение прямой: а) проходящей через точку $A\ (1,\ -1)$ и имеющей наклон -1; б) проходящей через точку $B(2,3)$ и имеющей наклон 2.

1.3.4. Докажите, что уравнение каждой прямой ненулевого (и не бесконечного) наклона (прямой, не параллельной осям координат) может быть записано в виде $\frac{x}{a}+\frac{y}{b}=1$ (уравнение прямой в отрезках на осях). Какой геометрический смысл имеют фигурирующие в этом уравнении величины a и b?

§ 4. Обратная пропорциональность и гипербола. Парабола

Напомним прежде всего понятие *прямой пропорциональности* двух величин. Зависимость

$$y=kx \qquad (1.4.1)$$

(геометрически она, как мы знаем, выражается прямой линией) означает, что y изменяется *пропорционально* x: во сколько раз увеличивается x, во столько же раз увеличится и y. Другими словами, для любых точек $A(x_1, y_1)$ и $B(x_2, y_2)$, координаты ко-

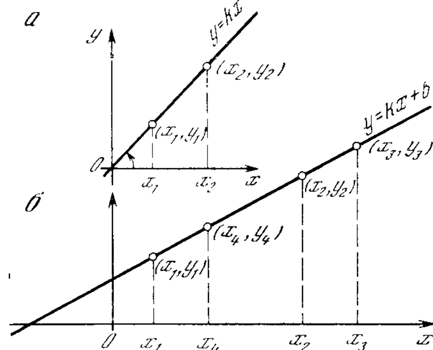


Рис. 1.4.1

торых удовлетворяют зависимости (1), всегда

$$\frac{y_2}{y_1}=\frac{x_2}{x_1}$$

(рис. 1, *а*). Говорят еще, что формула (1) выражает *прямую пропорциональность* между y и x (с *коэффициентом пропорциональности* k). Более общее уравнение

$$y=kx+b \qquad (1.4.1а)$$

характеризует пропорциональность (прямо пропорциональную зависимость) между приращениями x_2-x_1 независимой переменной и отвечающими им приращениями y_2-y_1 функции: для любых двух приращений x_2-x_1 и x_4-x_3 аргумента функции (1а) отвечающие им приращения самой функции будут пропорциональны приращениям аргументов (рис. 1, *б*):

$$\frac{y_2-y_1}{x_2-x_1}=\frac{y_4-y_3}{x_4-x_3}=k.$$

Другой часто встречающийся вид зависимости между y и x описывается формулой

$$y=\frac{k}{x}. \qquad (1.4.2)$$

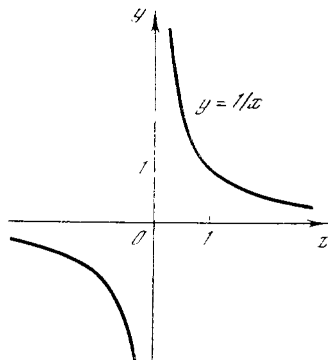
Такая зависимость называется *обратной пропорциональностью* (с *коэффициентом* k) между y и x: здесь если точки $A(x_1, y_1)$ и $B(x_2, y_2)$ удовлетворяют уравнению (2), то x_2 во столько раз больше x_1, во сколько y_2 меньше y_1:

$$\frac{x_2}{x_1}=\frac{y_1}{y_2}$$

$\left(\text{ибо } y_2:y_1=\frac{k}{x_2}:\frac{k}{x_1}=x_1:x_2\right)$.

Заметим еще, что отношения прямой и обратной пропорциональности двух величин *взаимны*: если y прямо (обратно) пропорционально x, то и x прямо (обратно) пропорционально y. Однако в то время как коэффициент прямой пропорциональности, связывающей x с y, обратен коэффициенту (прямой же) пропорциональности между y и x $\left(\text{если } y=kx, \text{ то } x=\frac{1}{k}y\right)$, в случае обратной пропорциональности величин y и x коэффициенты (обратной) пропорциональности, связывающие y с x и x с y, одинаковы $\left(\text{если } y=\frac{k}{x}, \text{ то и } x=\frac{k}{y}\right)$.

Рис. 1.4.2



Линия, уравнение которой имеет вид (2), называется *гиперболой* [1]. На рис. 2 изображена «единичная гипербола»

$$y=\frac{1}{x}, \qquad (1.4.2а)$$

а ниже приведены значения y при некоторых значениях x:

x	-1	$-0{,}1$	$-0{,}01$	$-0{,}001$	$0{,}001$	$0{,}01$	$0{,}1$	1
y	-1	-10	-100	-1000	1000	100	10	1

[1] Или *равнобочной гиперболой*, поскольку «гиперболами» математики называют линии, несколько более общие, чем кривая (2). Мы, однако, нигде не встретимся с этими «общими гиперболами» и потому прилагательное «равнобочная» будем отбрасывать.

Особенность гиперболы состоит в том, что при x малом положительном y очень велико, а при малых отрицательных x значения y также отрицательны и по абсолютной величине (по модулю) очень велики.

Эту особенность кривой (2а) (или общей кривой (2)) иногда выражают, говоря, что при $x=0$ мы имеем значение $y=\pm\infty$, где знак «+» или «—» зависит от того, с какой стороны мы подходим к значению $x=0$. Точный смысл записи «$y=\pm\infty$ при $x=0$» заключается в том, что при достаточно малом x величина y становится *сколь угодно большой* по абсолютной величине, причем в зависимости от знака x она может быть положительной или отрицательной (символ ∞ читается как «бесконечность»; разумеется, это не число). При этом чем меньшее по абсолютной величине значение x мы выбираем, тем выше (или при $x<0$ тем ниже) располагается соответствующая точка гиперболы: обе ветви гиперболы (2а) «уходят в бесконечность» (положительную или отрицательную) в направлении оси ординат, неограниченно приближаясь к этой оси (ведь абсцисса x может быть сколь угодно малой по абсолютной величине!), но никогда ее не достигая. Аналогично при неограниченном росте x по абсолютной величине (когда x «стремится к $+\infty$» или «стремится к $-\infty$»; еще раз напоминаем, что символ ∞ не отвечает никакому числу!) величина $y=\frac{1}{x}$ становится сколь угодно малой по абсолютной величине: гипербола «стремится к бесконечности в направлении оси x», неограниченно приближаясь к этой оси и никогда ее не достигая (причем две ветви гиперболы приближаются к оси x с разных сторон). Это свойство гиперболы выражают, говоря, что прямые $x=0$ (ось ординат) и $y=0$ (ось абсцисс) являются *асимптотами* гиперболы (от греческого слова asymptotos — несовпадающий).

Как видно из рис. 2, гипербола (2а) состоит из двух частей, или «ветвей», отвечающих значениям $x>0$ и значениям $x<0$; эти части разъединены, не связаны между собой.

Произвольная линия (2) получается из линии (2а) простым преобразованием. Пусть $A_0(x_0, y_0)$ — некоторая точка линии (2а), т. е. $y_0=\frac{1}{x_0}$ (см. рис. 3, на котором линия (2а) изображена штрихами). В таком случае тому же значению $x=x_0$ отвечает на линии (2), где мы считаем, что $k>0$, точка A с координатами x_0 и $y=\frac{k}{x_0}=k\frac{1}{x_0}$, т. е. $x=x_0$, $y=ky_0$.

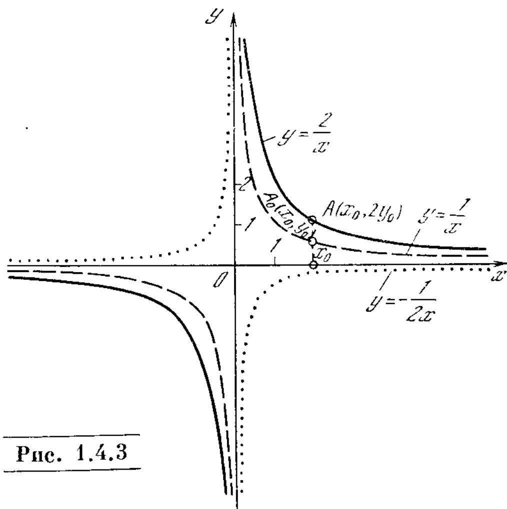


Рис. 1.4.3

Эта точка получается из точки A_0 «растяжением» вдоль оси y (или «растяжением» от оси x) с коэффициентом растяжения k (т. е. k-кратным увеличением всех вертикальных размеров) — она в k раз больше удалена от оси x, чем точка A_0.

Здесь мы считаем, что $k>1$; при k положительном, но меньшем единицы линия (2) будет получаться из линии (2а) «сжатием» вдоль оси y (или «сжатием» к оси x), ибо каждая точка будет ближе к оси x, чем соответствующая точка линии (2а) (ведь отношение $y:y_0=AP:A_0P=k$ в этом случае будет меньше единицы) [2].

На рис. 3 изображены линии $y=\frac{1}{x}$ (штриховая) и $y=\frac{2}{x}$ (сплошная). Точками на том же чертеже изображен график функции $y=\frac{-1/2}{x}$, отвечающей отрицательному значению $k=-1/2$; эта функция отрицательна при положительном x и положительна при отрицательном x.

---

[2] По поводу связи между линиями (2) и (2а) см. также упр. 3.

Прямая и обратная пропорциональности очень часто встречаются в физических законах. Так, *закон Ома* (1.3) означает, что сила тока i изменяется прямо пропорционально разности потенциалов u и обратно пропорционально сопротивлению R. При заданном R сила тока i пропорциональна u (с коэффициентом пропорциональности $\frac{1}{R}$); напротив, при заданном u величина i обратно пропорционвльна R: здесь $i = \frac{k}{R}$, где роль коэффициента k играет напряжение u. Аналогично этому простая зависимость $s = vt$ между путем s, скоростью v равномерного движения и временем t показывает, что время $t = \frac{s}{v}$ прямо пропорционально пути s и обратно пропорционально скорости v; давление газа в силу закона Бойля—Мариотта обратно пропорционально его объему и т. д. При этом формулы $s = vt$ и $t = \frac{s}{v}$ указывают, что при равномерном движении путь s пропорционален времени

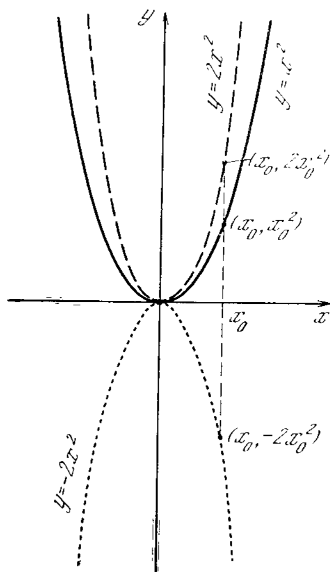


Рис. 1.4.4

движения t с коэффициентом пропорциональности v, в то время как время t пропорционально пути s с коэффициентом $\frac{1}{v}$: если $v = 20$ м/с, то коэффициенты рассматриваемых пропорциональностей равны 20 м/с и 0,05 с/м.

Иначе обстоит дело, скажем, в случае закона Бойля—Мариотта: PV=const. Если температура равна 0° С, а объем V и давление P измеряются в единицах м^3 и Па (паскаль, или ньютон на квадратный метр: 1 Па=1 Н/м^2=0,00098 кгс/см^2), то PV=0,8737 Н·м (ньютоно-метр); таким образом, P обратно пропорционально V, а V обратно пропорционально P, причем в обоих случаях мы имеем о д и н и т о т ж е коэффициент пропорциональности 0,8737 Н·м.

Эти примеры хорошо поясняют разницу между коэффициентами прямой пропорциональности $y=kx$ и обратной пропорциональности $y = \frac{k_1}{x}$. Если x измеряется в единицах e_1, а y — в единицах e_2, то k имеет размерность $\frac{e_2}{e_1}$, и поэтому k никак не может служить коэффициентом k' пропорциональности $x = k'y$, ибо последний должен измеряться в единицах $\frac{e_1}{e_2}$. Напротив, k_1 имеет размерность $e_1 e_2$, откуда вытекает, что эта величина может совпадать (и на самом деле совпадает) с коэффициентом k'_1 пропорциональности $x = \frac{k'_1}{y}$.

Рассмотрим теперь *квадратичную* функцию $y=ax^2+bx+c$. Начнем с простейшего случая: пусть

$$y = ax^2, \tag{1.4.3}$$

где коэффициент a может быть любым. Примем сначала для простоты, что a=1, т. е. рассмотрим линию [3]

$$y = x^2. \tag{1.4.3а}$$

Каковы общие свойства этой функции?

1. Всегда $y > 0$, как при $x > 0$, так и при $x < 0$; лишь при x=0 мы имеем y=0. Значит, вся кривая расположена в ы ш е оси x, только в начале координат она касается оси x.

2. Наименьшее значение (*минимум*) y достигается при x=0; он равен 0 (на графике минимум изображается самой низкой точкой кривой).

3. Двум одинаковым по абсолютной величине и противоположным по знаку значениям x отвечают одинаковые (как по знаку, так и по абсолютной величине) значения y. Значит, кривая *симметрична относительно оси* y.

Кривая (3а) изображена на рис. 4 сплошной линией. Эта кривая называется *параболой*, — и название «па-

[3] Этот случай в определенном смысле можно считать общим (см. § 8).

рабола» сохраняется за всеми линиями (3), отвечающими л ю б о м у значению a (и даже за более общим классом линий, как мы поясним ниже). При каждом положительном a линия (3) обладает теми же свойствами 1—3, что и линия (3а). В самом деле, переход от уравнения (3а) к уравнению (3) во всем аналогичен переходу от «единичной гиперболы» (2а) к «общей гиперболе» (2) (с положительным коэффициентом k): линия (3) получается из (3а) растяжением в a раз вдоль оси y (см. рис. 4, на котором изображены параболы $y=x^2$ и $y=2x^2$).

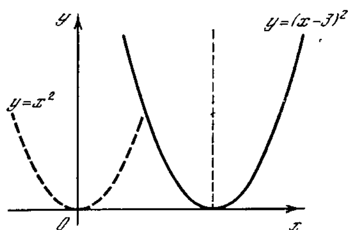


Рис. 1.4.5

Как будет выглядеть парабола при $a<0$? Рассмотрим пример $a=-2$, т. е. кривую $y=-2x^2$. Кривая изображена точками на том же рис. 4. Свойства этой кривой таковы:

1. $y<0$ при любых $x\neq 0$, т. е. кривая вся лежит н и ж е оси x и касается оси x в начале координат.

2. Функция y достигает наибольшего значения — *максимума* — при $x=0$. Этот максимум равен 0. (На графике максимум изображается верхней точкой кривой.)

3. Так же как и в случае положительного a, кривая симметрична относительно оси y.

Рассмотрим теперь еще более общее уравнение

$$y=a(x-n)^2. \qquad (1.4.4)$$

Пусть, например, $a=1$, $n=3$. Соответствующая кривая изображена на рис. 5. Это та же парабола $y=x^2$, но сдвинутая вправо вдоль оси x на три единицы.

Это простое обстоятельство обычно усваивается с трудом. Если дана функция $y=f(x)$ и мы сравниваем с ней функцию $y=f(x-n)$, то график второй функции смещен относительно первого графика на n единиц в п р а в о. При этом подразумевается, что в обоих случаях f — это одна и та же функция; так, в нашем примере, буква f означает операцию возведения в квадрат аргумента функции, т. е.

$$f(x)=x^2, \quad f(-x)=(-x)^2=x^2,$$

$$f(x-2)=(x-2)^2, \quad f(x-n)=(x-n)^2,$$

$$f(x^2)=(x^2)^2=x^4 \text{ и т. д.}$$

Почему же график смещается вправо? Поясним это подробнее. Пусть на графике функции $y=f(x)$ есть какая-то характерная точка $x=x_0$ (своего рода «зарубка») — в этой точке график функции может иметь излом, или функция может достигать своего максимума, или просто принимать какое-то фиксированное значение y_0 (рис. 6). Тогда на графике новой функции $y_1=f(x-n)$ то же значение или тот же излом появляется, когда аргумент функции f равен тому же значению x_0, что и ранее, т. е. при $x-n=x_0$. Но это значит, что координаты «зарубки» теперь таковы: $x=x_0+n$, $y_0=f(x_0)$. Отсюда видно, что любая «зарубка» вместе со всем графиком перемещается вправо: точке (x_0, y_0) первого графика отвечает точка (x_0+n, y_0) второго графика (см. рис. 6, а также сплошную и пунктирную параболы на рис. 5, где в качестве «зарубки» удобно выбрать точку $x_0=0$, $y_0=f(x_0)=0^2=0$).

Хотя эти соображения и очень элементарны, но весьма важно твердо их усвоить, не просто выучить, а осознать.

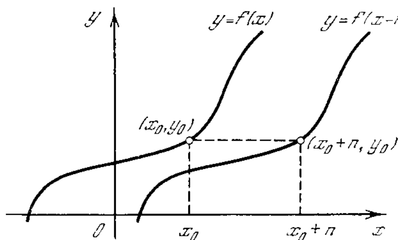


Рис. 1.4.6

Первое побуждение, которое возникает у многих учащихся, — ответить, что при переходе от уравнения $y=x^2$ к уравнению $y=(x-3)^2$ кривая сместится влево: ведь из величины x мы вычли 3. Надо не пожалеть времени на разбор примеров, демонстрирующих ошибочность этого.

Теперь можно сформулировать общие правила.

1. Кривая $y=a(x-n)^2$ имеет вертикальную ось симметрии, которой является прямая $x=n$.

2. Эта кривая при $a>0$ лежит выше оси x и достигает минимума $y=0$ при $x=n$; если же $a<0$, то кривая лежит ниже оси x и достигает максимума $y=0$ при $x=n$.

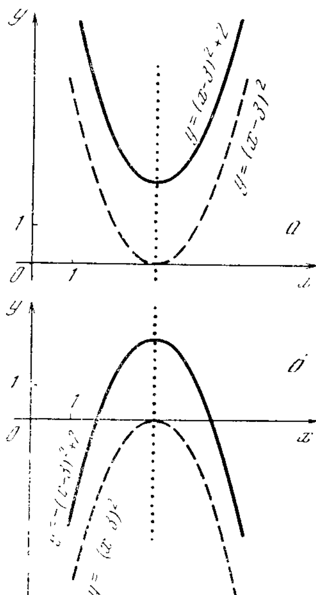


Рис. 1.4.7

Наконец, есть еще одно видоизменение уравнения, которое не меняет формы кривой. Рассмотрим функцию

$$y=a(x-n)^2+m. \qquad (1.4.5)$$

Очевидно, что соответствующая кривая (это тоже парабола!) отличается от предыдущей (в уравнении которой отсутствовало m) только сдвигом на m в вертикальном направлении. Положение же оси симметрии кривой не меняется: при $a>0$ функция достигает минимума при $x=n$, причем значение функции в точке минимума равно $y=m$ (ведь вместе со всей кривой сместился на m и минимум). При $a<0$ точка $x=n$, $y=m$ есть точка максимума кривой.

Приведем два примера (рис. 7): $y=(x-3)^2+2$, $y=-(x-3)^2+2$. Оси симметрии парабол на рис. 7 изображены точками; точка минимума на рис. 7, *а* и точка максимума на рис. 7, *б* лежат на пересечении кривой и ее оси симметрии.

Раскрывая в выражении $y=a(x-n)^2+m$ скобки, запишем

$$y=ax^2-2anx+an^2+m. \qquad (1.4.5а)$$

В правой части (5а) стоит многочлен второй степени. Многочлен второй степени в общем виде записывается так:

$$y=ax^2+bx+c. \qquad (1.4.6)$$

Последнюю формулу можно преобразовать так:

$$y=a\left(x^2+\frac{b}{a}x+\frac{c}{a}\right)=a\left(x^2+\frac{b}{a}x+\frac{b^2}{4a^2}\right)+a\left(\frac{c}{a}-\frac{b^2}{4a^2}\right)=a\left(x+\frac{b}{2a}\right)^2+\left(c-\frac{b^2}{4a}\right).$$

Следовательно,

$$ax^2+bx+c=a\left(x+\frac{b}{2a}\right)^2+\left(c-\frac{b^2}{4a}\right),$$

откуда и вытекает, что линия (6) — это тоже парабола с осью симметрии $x=-\frac{b}{2a}$ и точкой минимума или максимума $\left(-\frac{b}{2a},\ c-\frac{b^2}{4a}\right)$.

С помощью графика параболы можно разобраться и в решении *квадратного уравнения* для различных случаев, которые при этом возникают. К решению квадратного уравнения

$$ax^2+bx+c=0$$

можно подойти так: рассмотрим линию

$$y=ax^2+bx+c=a\left(x+\frac{b}{2a}\right)^2+\left(c-\frac{b^2}{4a}\right)$$

и найдем точки пересечения этой линии с осью x. В этих точках $y=0$, поэтому значения x, соответствующие точкам пересечения, и будут корнями квадратного уравнения.

Но мы знаем, что кривая $y=ax^2+bx+c$ есть парабола. Нам известно, что у этой параболы есть ось симметрии — вертикаль $x=-\frac{b}{2a}$, что при $a>0$ парабола имеет точку минимума на оси симметрии, ордината которого равна $y=c-\frac{b^2}{4a}$ (мы смотрим на вторую часть последней формулы, имеющую привычный вид $a(x-n)^2+m$). «Рога» параболы при $a>0$ обращены вверх.

Ясно, что если минимум лежит выше оси x, то парабола не пересекает ось x (рис. 8, кривая *1*). Значит, при

$$a>0, \quad c-\frac{b^2}{4a}>0$$

квадратное уравнение не имеет корней[4]. Если же минимум лежит ниже оси x, а «рога» параболы поднимаются вверх, то парабола обязательно пересечет ось x в двух точках, которые симметричны относительно прямой $x = n = -\frac{b}{2a}$ — оси симметрии параболы (рис. 8, кривая *2*). Значит, при

$$a > 0, \quad c - \frac{b^2}{4a} < 0$$

уравнение имеет два корня x_1 и x_2, изображенных на рис. 8.

Наконец, возможен и промежуточный случай, когда парабола касается оси x (рис. 8, кривая *3*). Этот случай достигается при

$$c - \frac{b^2}{4a} = 0.$$

Если постепенно переходить от кривой *2* к кривой *3*, поднимая параболу, то два корня x_1 и x_2 будут сближаться, пока в момент касания параболы с осью абсцисс они не сольются в одну точку. Поэтому в случае $c - \frac{b^2}{4a} = 0$ говорят не об одном корне, а о двух равных (слившихся) корнях уравнения.

Таким же способом рассматривается и случай $a < 0$, когда кривая имеет максимум, а «рога» ее обращены вниз. Предлагаем читателю самому нарисовать кривые и проверить, что при $a < 0$, $c - \frac{b^2}{4a} < 0$ уравнение не имеет корней; при $a < 0$, $c - \frac{b^2}{4a} > 0$ оно имеет два корня; при $a < 0$, $c - \frac{b^2}{4a} = 0$ мы приходим к случаю касания, когда уравнение имеет два одинаковых корня.

Обычная формула для корней квадратного уравнения такова:

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Уравнение имеет два вещественных корня в том случае, когда существует корень квадратный из числа $b^2 - 4ac$, т. е. когда $b^2 - 4ac > 0$. Запишем последнее выражение так:

$$b^2 - 4ac = -4a\left(c - \frac{b^2}{4a}\right).$$

Отсюда видно, что условие $b^2 - 4ac > 0$ выполняется в двух случаях:

1) $a > 0$, $c - \frac{b^2}{4a} < 0$; 2) $a < 0$, $c - \frac{b^2}{4a} > 0$.

Это и есть те два случая существования двух корней, которые мы получили выше из рассмотрения кривых.

К вопросу о решении квадратных уравнений можно подойти также и несколько по-другому. Разделим все члены уравнения

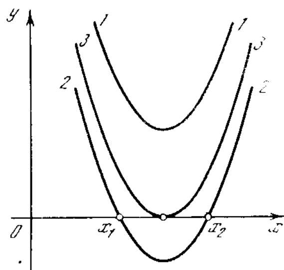


Рис. 1.4.8

$ax^2 + bx + c = 0$ (где, разумеется, $a \neq 0$) на a; мы придем тогда к равенству

$$x^2 + px + q = 0, \quad \text{или} \quad x^2 = -px - q, \qquad (1.4.7)$$

где $p = \frac{b}{a}$, $q = \frac{c}{a}$. Изобразим на рисунке графики двух функций:

$$y_1 = x^2 \qquad (1.4.7а)$$

(это «единичная» парабола (3а)) и

$$y_2 = -px - q \qquad (1.4.7б)$$

(прямая линия). Ясно, что при тех значениях x, которые удовлетворяют уравнению (7), мы имеем $y_1 = y_2$, т. е. эти значения x отвечают точкам пересечения параболы (7а) и прямой (7б)! Но параболу (7а) мы можем нарисовать (скажем, на миллиметровой бумаге) с большой точностью; тогда для решения любого квадратного уравнения достаточно будет провести на том же листе бумаги прямую (7б) (скажем, найдя две какие-либо точки этой прямой и воспользовавшись затем линейкой), после чего мы сможем «прочитать» на чертеже значения корней квадратного уравнения. При этом прямая (7б) может или пересечь параболу (7а) в двух точках (как прямая *1* на рис. 9), или коснуться ее в единственной точке (иметь с параболой (7а) две совпадающие точки пересечения (прямая *2* на рис. 9)), или же может вовсе не иметь общих точек с (7а) (прямая *3* на рис. 9). Этому отвечают случаи, когда уравнение (7) имеет два (различ-

[4] Здесь под словом «корень» понимается *вещественный* корень уравнения; о так называемых *комплексных* числах, которым посвящены главы 14, 15 третьей части книги, мы пока просто забываем.

ных) корня, два совпадающих корня (т. е. один корень), ни одного корня соответственно.

Из сказанного можно сделать также следующее заключение. Мы знаем, что уравнение $ax^2+bx+c=0$ имеет единственный корень в том и только в том случае, когда $c-\frac{b^2}{4a}=0$; поэтому уравнение $x^2+px+q=0$ (случай $a=1$)

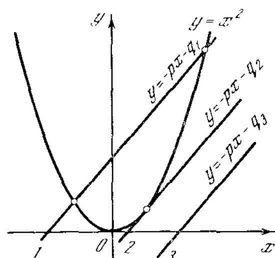


Рис. 1.4.9

имеет единственный корень тогда и только тогда, когда

$$q-\frac{p^2}{4}=0. \tag{1.4.8}$$

Таким образом, (8) есть условие касания прямой (7б) и параболы (7а), т. е. прямая $y=kx+b$ касается параболы $y=x^2$ в том и только в том случае, когда [5]

$$k^2+4b=0. \tag{1.4.8a}$$

В гл. 2 мы выведем то же условие (8а) касания прямой и параболы другим способом.

Заметим, наконец, что в зависимости от знака коэффициента a при x^2 парабола (6) будет обращена выпуклостью вниз (случай $a>0$) или выпуклостью вверх (случай $a<0$). Это свойство не зависит от значений и знаков b и c в уравнении (6) параболы.

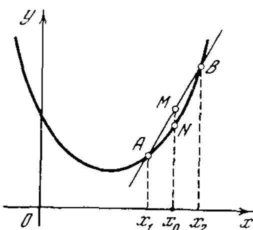


Рис. 1.4.10

Точное определение выпуклости таково: выберем на кривой какие-либо две точки A (x_1, y_1) и B (x_2, y_2) и проведем через них прямую (рис. 10). Если часть кривой, расположенная между этими точками, лежит ниже соответствующего отрезка прямой (ниже хорды AB кривой), то говорят, что кривая обращена *выпуклостью вниз*. Если часть кривой, расположенная между точками, лежит выше хорды AB, то кривая обращена *выпуклостью вверх*.

Характер выпуклости параболы легко усмотреть из чертежа, но можно установить и алгебраически. Для этого выберем два (какие угодно!) значения x_1 и x_2 абсциссы x; им отвечают точки A (x_1, y_1) и B (x_2, y_2) параболы, где $y_1=ax_1^2+bx_1+c$, $y_2=ax_2^2+bx_2+c$.

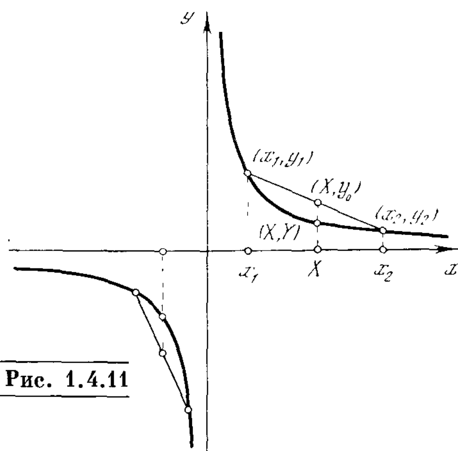


Рис. 1.4.11

Найдем теперь координаты середины M хорды AB параболы (см. рис. 10). Можно показать геометрически, что если $AM=MB$, то координаты точки M (x_0, y_0) суть средние арифметические координат A и B:

$$x_0=\frac{x_1+x_2}{2} \quad \text{и} \quad y_0=\frac{y_1+y_2}{2}.$$

(Это следует из того, что на рис. 10 средняя линия Mx_0 трапеции ABx_2x_1 равна полусумме оснований Ax_1 и Bx_2 трапеции.) Вычислим далее координаты точки N (X, Y) параболы, отвечающей тому же значению $X=x_0=\frac{x_1+x_2}{2}$ абсциссы. Здесь

$$Y=aX^2+bX+c=$$
$$=a\left(\frac{x_1+x_2}{2}\right)^2+b\frac{x_1+x_2}{2}+c.$$

Читатель легко убедится, что

$$Y-y_0=a\left(\frac{x_1+x_2}{2}\right)^2-\left(a\frac{x_1^2}{2}+a\frac{x_2^2}{2}\right)=$$
$$=-a\left(\frac{x_1-x_2}{2}\right)^2$$

(ибо члены с коэффициентами b и c взаимно уничтожаются). Но вели-

[5] Ясно, что условие (8) переходит в (8а), если заменить в (8) $-p$ на k и $-q$ на b.

чина $\left(\frac{x_1 - x_2}{2}\right)^2$ положительна при любых x_1, x_2. Следовательно, при $a > 0$ имеем $Y < y_0$ — точка на параболе лежит ниже соответствующей (имеющей ту же абсциссу X) точки прямой, т. е. парабола обращена выпуклостью вниз; при $a < 0$, напротив, $y_0 < Y$, т. е. парабола обращена выпуклостью вверх.

Гипербола $y = \frac{k}{x}$ (где мы считаем, $k > 0$) состоит из двух частей, или «ветвей». Нетрудно видеть, что часть («ветвь») гиперболы, отвечающая положительным значениям x, обращена выпуклостью вниз, а вторая ветвь гиперболы обращена выпуклостью вверх (рис. 11; см. также упр. 4).

Упражнения

1.4.1. Изобразите графики функций $y = \frac{3}{x}$, $y = -\frac{0{,}5}{x}$, $y = \frac{1}{x} + 3$.

1.4.2. Изобразите графики функций $y = x^2 - 2x + 2$, $y = 2x^2 + 4x$.

1.4.3. Покажите, что гипербола $y = \frac{k}{x}$, где $k > 0$, может быть получена из гиперболы $y = \frac{1}{x}$:

а) растяжением от начала координат O (гомотетией) с коэффициентом $\sqrt{k}$ (при $k < 1$ уместнее говорить не о растяжении, а о сжатии); это значит, что если M' и M — соответственно точки линий $y = \frac{k}{x}$ и $y = \frac{1}{x}$, лежащие на одной прямой с началом координат O, то $OM' : OM = \sqrt{k}$;

б) растяжением в направлении оси x (или растяжением от оси y) с коэффициентом k (и тут при $k < 1$ уместнее говорить не о «растяжении», а о «сжатии»).

1.4.4. Докажите алгебраически, что при $k > 0$ гипербола $y = \frac{k}{x}$ при $x > 0$ обращена выпуклостью вниз, а при $x < 0$ — выпуклостью вверх.

§ 5. Параболы и гиперболы высших порядков. Полукубическая парабола

Кривую, являющуюся графиком функции

$$y = ax^n \tag{1.5.1}$$

(где n — натуральное число), часто называют *параболой порядка n*. Так, на рис. 1 изображены *парабола 3-го порядка* (кубическая парабола)

$$y = x^3 \tag{1.5.2}$$

и *парабола 4-го порядка*

$$y = x^4. \tag{1.5.3}$$

Мы видим, что парабола 4-го порядка (3) по виду напоминает обычную (ква-

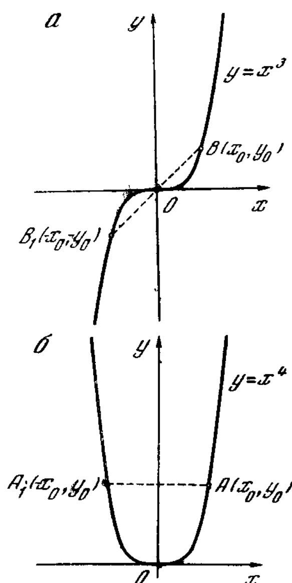


Рис. 1.5.1

дратичную) параболу: она имеет ось симметрии, которой служит ось y; в точке $x=0$ имеет минимум; как и парабола $y=x^2$, она касается в начале координат оси x, не пересекая ее.

Парабола 3-го порядка $y=x^3$ обладает другими свойствами. Эта кривая не имеет ни максимума, ни минимума: здесь увеличение x в с е г д а влечет за собой и увеличение y, т. е. при движении точки по линии слева направо она все время поднимается. Говорят, что функция (2) при всех x *возрастает* (здесь имеется в виду рост функции при увеличении ее аргумента). Кривая (2) не имеет оси симметрии: значения y в этом случае отрицательны при отрицательных x и положительны при положительных x. Зато у нее есть *центр симметрии*, которым служит начало координат O (0, 0). В самом деле, в случае функции (3) равным по абсолютной величине, но противоположным по знаку значениям x_0 и $-x_0$ аргумента отвечают о д и н а к о в ы е значения $y_0=x_0^4$ самой функции, т. е. каждой точке A (x_0, y_0) линии (3) отвечает

симметричная относительно оси y точка A_1 $(-x_0,\ y_0)$ (рис. 1, *б*). В случае же функции (2) равным по абсолютной величине и противоположным по знаку значениям x_0 и $-x_0$ аргумента отвечают противоположные

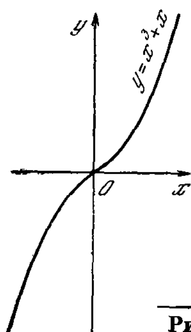


Рис. 1.5.2

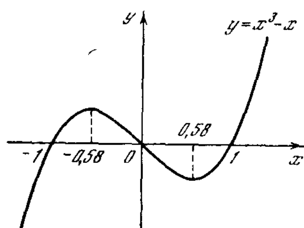


Рис. 1.5.3

значения $y_0=x_0^3$ и $-y_0=-x_0^3$ функции, так что здесь каждой точке B $(x_0,\ y_0)$ параболы отвечает *симметричная* относительно точки O точка B_1 $(-x_0,\ -y_0)$, такая, что O является серединой отрезка BB_1 (рис. 1, *а*).

На рис. 2 изображена кривая

$$y=x^3+x. \tag{1.5.4}$$

Эта кривая также характеризуется тем, что на любом ее участке при увеличении x увеличивается и y: функция (4), подобно функции $y=x^3$, растет при всех x. Кривая (4) не имеет ни максимумов, ни минимумов. Ясно, что подобная кривая пересекает ось абсцисс только один раз, а именно при $x=0$.

На рис. 3 показана кривая

$$y=x^3-x. \tag{1.5.5}$$

Как видно из графика, на этой кривой есть два участка, где y возрастает с увеличением x: при отрицательных $x<-0{,}58$ и при положительных $x>0{,}58$. Между ними — на отрезке $-0{,}58<x<+0{,}58$ — функция является убывающей: с ростом x величина y убывает.

Функция (5) имеет *максимум* при $x\approx-0{,}58$, $y\approx0{,}38$ [1]. Однако слово «максимум» вовсе не означает в данном случае, что $y\approx0{,}38$ есть вообще наибольшее возможное значение y, заданного выражением (5); ведь ясно, что при больших положительных значениях x величина y принимает сколь угодно большие значения. Так чем же выделяется точка максимума ($x\approx-0{,}58$, $y\approx0{,}38$) функции (5)?

Как видно из графика, в этой точке y больше, чем в соседних точках. Точка максимума отделяет участок, где функция является растущей (слева от максимума), от участка, где функция убывает (справа от максимума). Такой максимум называют *локальным* (местным), потому что величина y в этой точке больше значений y в других точках лишь для x, соседних со значением $x_{\max}$ (в нашем случае — со значением $x\approx-0{,}58$), не слишком далеких от этого значения. Аналогично в точке $x\approx+0{,}58$ функция имеет локальный минимум $y\approx-0{,}38$.

На рис. 4 приведены еще два примера кривых, изображающих многочлены третьей степени. Кубическое уравнение, полученное приравниванием нулю

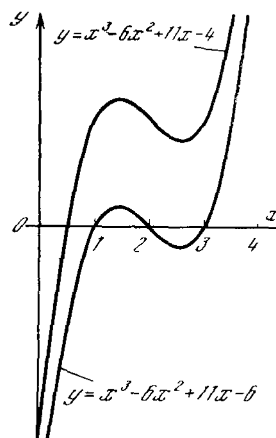


Рис. 1.5.4

многочлена, в случае верхней кривой имеет одно (вещественное) решение (один корень): $x\approx0{,}48$, а в случае нижней кривой — три корня: $x_1=1$, $x_2=2$, $x_3=3$. Легко убедиться, что кубическое уравнение всегда имеет *не менее одного* (вещественного) *корня*; для этого читателю предлагается подумать о ходе кривой $y=ax^3+bx^2+cx+d$ при очень больших по абсолютной величине положительных и отрицательных значениях x.

[1] Ниже, в § 2.6, мы научимся вычислять точки максимума кривой, заданной явной формулой. В частности, в случае формулы (5) окажется, что $x_{\max}=-1/\sqrt{3}\approx0{,}57735$, $y_{\max}=2/(3\sqrt{3})\approx0{,}3849$. (Индекс max при x — сокращение латинского слова maximum — *максимум*, читается: икс-максимум; индекс min расшифровывается как minimum — *минимум*.)

График кубической параболы (2) может быть использован для (приближенного) решения любых *уравнений 3-й степени*. Запишем общее уравнение 3-й степени (кубическое уравнение) так:

$$x^3 + 3ax^2 + bx + c = 0 \qquad (1.5.6)$$

(нам будет удобно обозначать коэффициент при x^2 через $3a$, а не просто через a). Преобразуем уравнение (6) следующим образом:

$$(x^3 + 3ax^2 + 3a^2x + a^3) + (b - 3a^2)\,x + {} + (c - a^3) = 0,$$

или

$$(x + a)^3 - (3a^2 - b)(x + a) - {} - [(a^3 - c) - a(3a^2 - b)] = 0,$$

или, наконец,

$$X^3 - KX - B = 0, \qquad (1.5.6а)$$

где $X = x + a$, $K = 3a^2 - b$, $B = ab - c - 2a^3$.

Ясно, что решить уравнение (6) — это то же самое, что решить уравнение (6а), которое можно переписать еще так:

$$X^3 = KX + B. \qquad (1.5.7)$$

Поэтому если мы имеем выполненный с хорошей точностью (скажем, на миллиметровой бумаге) график кубической параболы $y = X^3$, то нам достаточно лишь построить на том же чертеже прямую $y_1 = KX + B$ и найти точки (или одну точку) пересечения кубической параболы и прямой — так мы сумеем определить примерную величину корней уравнения (6а) (а значит, и уравнения (6)).

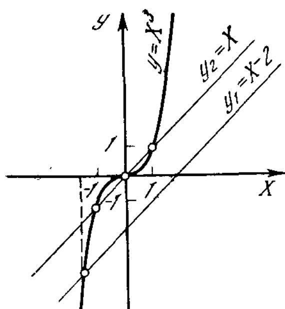


Рис. 1.5.5

Например, в случае уравнения $x^3 - 6x^2 + 11x - 4 = 0$ (см. рис. 4) мы имеем $a = -2$, $b = 11$, $c = -4$, откуда

$$K = 3a^2 - b = 3 \cdot 4 - 11 = 1,$$

$$B = ab - c - 2a^3 = -2 \cdot 11 + 4 - 2 \cdot (-8) = -2.$$

Таким образом, для решения нашего уравнения достаточно найти точки пересечения параболы $y = X^3$ и прямой $y_1 = X - 2$. Из рис. 5 мы получаем приближенное решение; $X \approx \approx -1{,}5$, откуда $x = X - a \approx -1{,}5 + 2 \approx 0{,}5$.

Аналогично в случае уравнения $x^3 - 6x^2 + 11x - 6 = 0$ (см. рис. 4) имеем $a = -2$, $b = 11$, $c = -6$, $K = 3a^2 - b = 1$, $B = ab - c - 2a^3 = 0$. Здесь, очевидно, точки пересечения кубической параболы $y = X^3$ и прямой $y_2 = X$ отвечают значениям X, равным -1, 0 и $+1$, откуда следует, что корни $x = X - a = X + 2$ исходного уравнения равны 1, 2 и 3 (см. рис. 5).

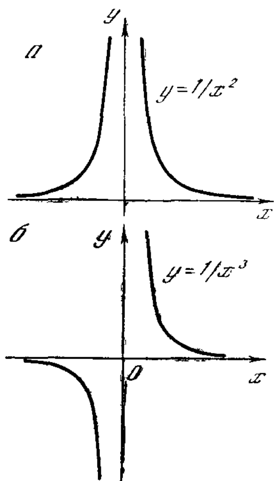


Рис. 1.5.6

Кривую, являющуюся графиком функции

$$y = \frac{k}{x^n} \quad (= kx^{-n}), \qquad (1.5.8)$$

где n — натуральное число, иногда называют *гиперболой* $(n+1)$*-го порядка*, или *гиперболой* n*-й степени*. Так, на рис. 6 изображены *гипербола* 2-й степени

$$y = \frac{1}{x^2} \qquad (1.5.9)$$

и *гипербола 3-й степени*

$$y = \frac{1}{x^3}. \qquad (1.5.10)$$

Легко понять, что кривая (10) по характеру напоминает обыкновенную гиперболу $\frac{1}{x}$: обе эти кривые состоят из двух (симметричных относительно начала O координат) ветвей: при $x = 0$ величина y здесь принимает значение $\pm\infty$ (ср. со сказанным в § 4 по поводу гиперболы). Напротив, гипербола 2-й степени (9) по виду отлична от обыкновенной гиперболы, поскольку здесь y положительно при всех (положительных и отрицательных) значениях x. Эта линия также состоит из двух ветвей, но симметричных уже не относительно точки — начала координат, а относительно прямой — оси y.

Для функции $y=\frac{1}{x^2}$ можно символически записать, что $y=+\infty$ при $x=0$ (это означает, что при достаточно малых x величина y становится сколь угодно большой).

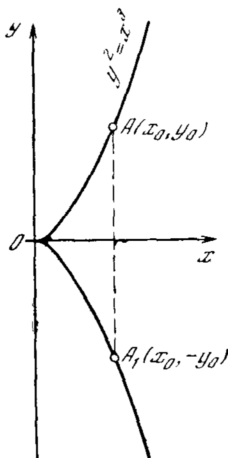


Рис. 1.5.7

График функции

$$y=\pm x^{3/2}=\pm\sqrt{x^3}, \tag{1.5.11}$$

точнее, кривую с уравнением

$$y^2=x^3 \tag{1.5.11a}$$

называют *полукубической параболой* (рис. 7). Так как отрицательным значениям x здесь не отвечают никакие значения y (при отрицательных x правая часть (11a) отрицательна, что невозможно, ибо y^2 не отрицательно ни при каких y), то вся линия расположена в правой полуплоскости. Так как каждому значению x отвечают два противоположных по знаку значения y (а именно $+\sqrt{x^3}$ и $-\sqrt{x^3}$), то наша линия симметрична относительно оси x: каждой ее точке A (x_0, y_0) отвечает симметричная относительно оси абсцисс точка A_1 $(x_0, -y_0)$. (Ниже, в § 2.5, мы строго докажем, что полукубическая парабола в начале координат O касается оси x.)

Рассмотренные примеры иллюстрируют ход всех *степенных функций*

$$y=x^n, \tag{1.5.12}$$

где показатель степени n может быть положительным и неположительным, по абсолютной величине большим или меньшим 1, целым или дробным.

Если $n>1$, то кривые (12) (которые мы будем рассматривать лишь при неотрицательных x, ибо при нецелом n само определение величины x^n, где $x<0$, вызывает известные трудности; так, например, величина $x^{1/2}=\sqrt{x}$ при $x<0$ не существует) ведут себя как (квадратичная) парабола (4.3a). Все они касаются оси x в начале координат O и проходят через точку Q (1,1), причем, чем больше n, тем теснее примыкает кривая (12) к оси Ox в окрестности начала координат и тем круче устремляется она вверх на подходе к точке Q. При $0<n<1$ кривые (12), напротив, в начале координат O касаются оси y, прижимаясь к ней тем сильнее, чем меньше n; они тоже проходят через точку Q (1, 1), причем, чем меньше n, тем резче поворачивают они к этой точке, отходя от начала координат O (см. рис. 8, *а*, на котором изображены кривые (12), отвечающие различным значениям $n>0$, а также прямая $y=x$, отвечающая значению $n=1$ и отделяющая друг от друга линии (12) с $n>1$ и с $n<1$).

Совсем по-другому ведут себя кривые (12) при $n<0$. Эти линии

$$y=x^{-m}=\frac{1}{x^m},\quad \text{где } m>0, \tag{1.5.12a}$$

также проходят через точку Q (1, 1); однако, в противоположность случаю $n>0$, они

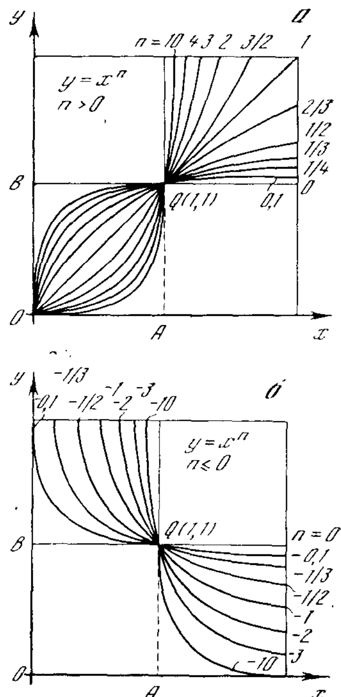


Рис. 1.5.8

вовсе не заходят в квадрат $OAQB$ с диагональю OQ, располагаясь целиком вне его. Оси x и y служат асимптотами этих кривых, причем чем больше n по абсолютной величине

(чем больше положительное число $m=-n$), тем быстрее стремятся они к (так никогда и не наступающему) слиянию с осью x и тем медленнее приближаются к оси y (см. рис. 8, *б*, где изображены кривые (12), отвечающие значениям $n=-1/10$; $-1/3$; $-1/2$; -1; -2; -3; -10; т. е. значениям $m=1/10$; $1/3$; $1/2$; 1; 2; 3; 10 в уравнении (12а), а также «предельная» линия $y=x^0=1$, отвечающая значению $n=0$).

Отметим еще, что через каждую точку $M(x, y)$ первой четверти координатной плоскости (точку $M(x, y)$, где $x, y > 0$), для которой $x \neq 1$, проходит единственная линия (12): этой линии отвечает значение $n > 0$, если разности $x-1$ и $y-1$ имеют одинаковые знаки (т. е. если $x > 1$ и $y > 1$ или $x < 1$ и $y < 1$), и значение $n < 0$, если разности $x-1$ и $y-1$ имеют разные знаки (если одно из чисел x, y больше 1, а второе — меньше 1). Что же касается точек $N(1, y)$, где $y > 0$, то через отличную от Q точку N не проходит ни одна кривая (12), а через Q проходит бесконечно много таких кривых (все).

По поводу поведения линий (12) при больших по абсолютной величине значениях n (по поводу «асимптотики» функций (12) при $n \to \infty$ или $n \to -\infty$, как говорят математики [2]) см. упр. 3.

Упражнения

1.5.1. Изобразите графики функций: а) $y=x^4+1$; б) $y=-x^4+1$; в) $y=x^4+x^2$.

1.5.2. Изобразите графики функций: а) $y=\frac{1}{x^3}+1$; б) $y=-\frac{1}{x^2}+4$.

1.5.3. Какой вид имеют графики функций: а) $y=x^{2n}$; б) $y=x^{2n+1}$; в) $y=\frac{1}{x^{2n}}$; г) $y=\frac{1}{x^{2n+1}}$ при очень больших (натуральных) n?

§ 6. Обратная функция. Графики взаимно-обратных функций

Задание y как функции x означает, что каждому x соответствует определенное значение y. Но эту зависимость можно и обратить: задавать y и по нему находить соответствующее значение x. Так, например, закон движения электропоезда от одного пункта до другого устанавливается зависимостью положения поезда z от времени t, где за координату z можно принять, например, расстояние поезда от начального пункта движения. Этот закон имеет вид функциональной зависимости $z=f(t)$; именно этой зависимостью z от t, заданной в виде расписания движения поезда, руководствуется его водитель.

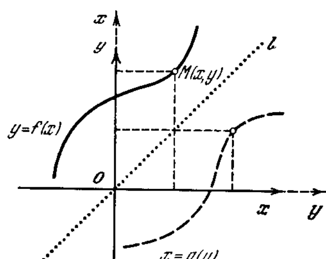


Рис. 1.6.1

Но для пассажира электропоезда больший интерес может представить иная форма зависимости между пройденным путем z и временем t, а именно зависимость времени от пути, указывающая, в какое время t_1 прибудет поезд в тот или иной пункт, определяемый координатой z_1. Эта зависимость $t=g(z)$ называется обратной по отношению к зависимости $z=f(t)$; другими словами, *g есть функция, обратная f*. Говорят также, что функции f и g являются взаимно-обратными.

Приведем несколько примеров (слева дана функция $y(x)$, а справа — обратная ей функциональная зависимость $x(y)$, где теперь, вопреки традициям, буквой y обозначен аргумент функции, а буквой x — сама функция):

$$
\begin{aligned}
&y=x+a, && x=y-a;\\
&y=3x+2, && x=\frac{1}{3}y-\frac{2}{3};\\
&y=1-x, && x=1-y; \qquad (1.6.1)\\
&y=x^2, && x=\pm\sqrt{y};\\
&y=x^3+1, && x=\sqrt[3]{y-1}.
\end{aligned}
$$

Нетрудно сообразить, как связаны графики взаимно-обратных функций f и g. Пусть мы имеем график функции $y=f(x)$ (рис. 1). Для того чтобы этот же чертеж мог служить графиком зависимости $x=g(y)$, надо посмотреть на него иначе: принять ось y за ось независимых переменных (ось абсцисс), а ось x — за ось ординат. Если желать,

---

[2] Слово «асимптотика» идет от уже знакомого нам термина «асимптота»; ведь утверждение о том, что функция $y=1/x$ имеет асимптоты $y=0$ или $x=0$, как раз и характеризует вид этой функции при малых (и при больших) по абсолютной величине x.

чтобы ось независимых переменных (абсцисс) по-прежнему была горизонтальной, а ось значений функции вертикальной, то этого можно добиться, повернув чертеж (наш график, а также оси координат) на 180° вокруг биссектрисы l I и III координатных углов.

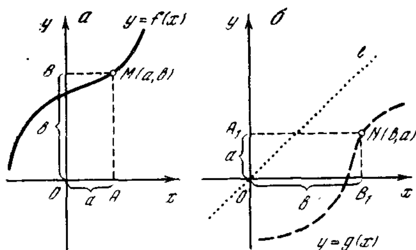


Рис. 1.6.2

После такого поворота ось y станет горизонтальной, а ось x — вертикальной. Результат такого поворота на рис. 1 изображен пунктиром; пунктиром же продолжены оси Ox и Oy, чтобы указать, что они поменяются местами.

На рис. 2 две части рис. 1 — сплошная и пунктирная — разъединены; при этом на рис. 2, *б* мы по-прежнему обозначаем абсциссы через x, а ординаты — через y. Из этих двух чертежей видно, что если на рис. 2, *а* значению $a=OA$ независимой переменной отвечает зна-

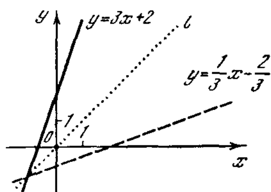


Рис. 1.6.3

чение $b=OB$ функции (где $b=f(a)$), то на рис. 2, *б*, изображающем график функции g, имеем $a=g(b)$. Поворот вокруг биссектрисы l на 180° переводит отрезки $OA=a$ и $OB=b$ рис. 2, *а* в отрезки $OA_1=a$ и $OB_1=b$ рис. 2, *б*. Таким образом, *графики взаимно-обратных функций симметричны относительно биссектрисы l I и III координатных углов*. Так, на рис. 3 изображены графики взаимно-обратных функций $y=3x+2$ и $y=\frac{1}{3}x-\frac{2}{3}$. В частности, если график функции симметричен относительно биссектрисы l, то функция совпадает со своей обратной. Таковы, скажем, функции $y=1-x$ или $y=\frac{1}{x}$ (см. рис. 4.2). И в самом деле, из $y=1-x$ следует, что $x=1-y$, а из $y=\frac{1}{x}$ вытекает, что $x=\frac{1}{y}$.

Обратимся теперь к рис. 4, изображающему графики функций $y=x^2$ (сплошная линия) и $y=\sqrt{x}$ (пунктир).

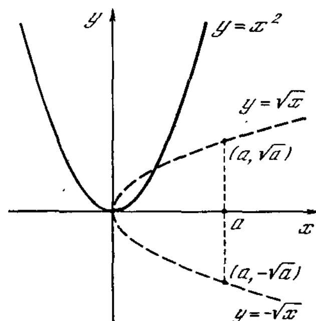


Рис. 1.6.4

Из этого рисунка видно, что функция $y=\pm\sqrt{x}$, обратная квадратичной функции $y=x^2$, является двузначной: каждому положительному значению x независимой переменной здесь отвечают два значения функции, а именно $y=\sqrt{x}$ и $y=-\sqrt{x}$. Для всех же отри-

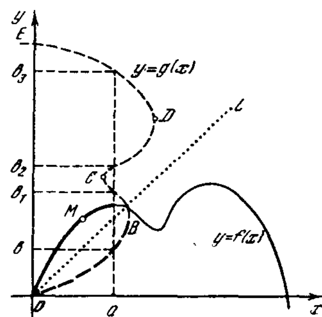


Рис. 1.6.5

цательных x отвечающее ему y не существует, ибо график функции $y=\pm\sqrt{x}$ целиком расположен в правой полуплоскости, отвечающей положительным x. Поэтому в рассматриваемом случае утверждение, что по заданному значению x можно сразу же указать отвечающее ему значение y обратной функции, будет неверным: для некоторых x (для $x<0$) отвечающее ему y н е с у щ е с т в у е т, для других же x (для $x>0$) мы имеем, пожалуй, даже слишком много отвечающих этому x значений y — сразу два.

Еще сложнее обстоит дело в случае функции $f(x)$, изображенной на рис. 5 (на том же рисунке пунктиром изображена обратная функция g). В этом слу-

чае, скажем, значению $x=a$ отвечают сразу четыре значения $g(x)$, а именно b, b_1, b_2 и b_3. Впрочем, само по себе это никак не может нас удивить: конечно, функция, определенная как обратная заданной $f(x)$, вовсе не обязана быть однозначной. Так, в примере, с которого мы начали параграф, пункты, которые проезжает электропоезд (они определяются расстоянием z от начального пункта O), разумеется, однозначно определяются по времени t: именно эта зависимость $z=z(t)$ дается расписанием движения поезда. Но если данный пригородный поезд совершает в течение суток ряд рейсов, то через определенный пункт своего маршрута он в течение этих суток будет проходить многократно. Таким образом, обратная зависимость времени от расстояния $t=t(z)$ будет *многозначной*: одному и тому же расстоянию z от начального пункта движения отвечает ряд моментов времени, когда электропоезд удален от O именно на выбранное расстояние z.

Так в чем же причина различия между функциями $y=3x+2$ (см. рис. 3; здесь обратная функция однозначна) и $y=x^2$ (см. рис. 4 — обратная функция двузначна) или, тем более, изображенной на рис. 5 функцией $f(x)$ (обратная функция многозначна)? Нетрудно понять, в чем тут дело. Процесс «обращения» функции — нахождения функции, обратной заданной, заключается в восстановлении значений абсцисс по отвечающим им ординатам — скажем, времени t по проходимому электропоездом пути z. Хорошо, если при этом каждому значению функции y отвечает единственное значение аргумента x, т. е. если обратная функция однозначна; однако этот «хороший» случай имеет место далеко не всегда. Но обратная функция всегда однозначна, если исходная однозначная функция $y=f(x)$ *монотонна*, т. е. если она все время растет или все время убывает. Здесь в процессе изменения аргумента x функция будет пробегать все новые значения y и каждому y будет отвечать свое значение x. Именно поэтому в случае функции $y=3x+2$ определение обратной функции не вызвало у нас никаких затруднений — ведь эта (линейная) функция монотонно растет. Так же обстоит дело в случае зависимости проходимого поездом пути от времени ($z=f(t)$), если поезд движется в одном направлении (без остановок). Но если направление изменения функции (скажем, направление движения поезда) меняется на обратное, например функция сначала убывает, как функция $y=x^2$, а затем, перейдя через точку минимума, начинает возрастать, то, возрастая, функция снова пробегает уже встречавшиеся ранее значения, — и поэтому мы не можем с полной определенностью сказать, какому именно x отвечает данное y, ибо таких x оказывается несколько (в случае функции $y=x^2$ — два). И уж совсем не подходит для определения обратной функции постоянная функция $y=a$ (см. рис. 3.6); ведь здесь y совсем не меняет своего значения и определить по нему x невозможно: функция $y=a$ совсем не имеет обратной.

В случае же иных функций дело обстоит вовсе не так уж плохо. Если мы хотим сконструировать однозначную функцию, обратную, скажем, функции $y=x^2$, нам надо лишь ограничиться монотонной частью этой функции, например, рассматривать ее при одних лишь положительных значениях аргумента x (соответствующая часть графика функции $y=x^2$ изображена на рис. 4 более жирной линией). Этому монотонному участку функции будет отвечать однозначная обратная функция. График такой однозначной ветви функции $y=\sqrt{x}$ изображен на рис. 4 более толстым пунктиром; именно эту функцию обычно и обозначают через $\sqrt{x}$ (иногда ее называют *арифметическим значением*[1] корня квадратного из x). Аналогично этому, если только данная точка $M(x_0, y_0)$ графика функции $y=f(x)$ не является точкой максимума или минимума, подобно, например, точке M на рис. 5, то в окрестности рассматриваемой точки M можно выделить участок монотонности[2] функ-

[1] Следует только иметь в виду, что величина $+\sqrt{x}$, где под корнем квадратным понимается арифметическое его значение, все же не является полностью обратной функции $y=x^2$: в точности обратна этой функции двузначная функция $y=\pm\sqrt{x}$.

[2] В соответствии с этим говорят о *локальной монотонности* функции $y=f(x)$ в точке M, т. е. о монотонности функции в некоторой окрестности точки M.

ции f. Этому участку отвечает некоторая однозначная ветвь обратной функции g (см. участки графиков функции $f(x)$ и $g(x)$, выделенные на рис. 5 более жирными линиями). Вообще многозначная функция вроде изображенной пунктиром на рис. 5 функции g обычно естественно распадается на однозначные ветви функции — в нашем случае такие ветви изображаются дугами OB, BC, CD и DE (см. рис. 5). Трудность здесь может состоять лишь в том, какую из этих ветвей целесообразно считать, так сказать, «главной»; впрочем с точки зрения математика, это, разумеется, совершенно безразлично.

Упражнения

1.6.1. Какие функции обратны функциям: а) $y=2x+4$; б) $y=x^2-2$; в) $y=x^3+3x^2+3x$; г) $y=x^4+2x^2$.

1.6.2. Использовав рис. 5.2, 5.3 и 5.6, *а*, *б*, нарисуйте графики функций, обратных функциям: а) $y=x^3+x$; б) $y=x^3-x$; в) $y=x^{-2}=\frac{1}{x^2}$; г) $y=x^{-3}=\frac{1}{x^3}$, выделив, если требуется, участки монотонности этих функций, позволяющие прийти к однозначной обратной функции.

1.6.3. Найдите функцию, обратную: а) произвольной линейной функции $y=ax+b$; б) произвольной квадратичной функции $y=ax^2+bx+c$; в) функции $y=\frac{ax+b}{cx+d}$.

1.6.4. Пусть $f(x)$ и $g(x)$ — взаимно-обратные функции, где функция f определена на участке $a \leqslant x \leqslant b$, а функция g — на участке $\alpha \leqslant x \leqslant \beta$, $f(a)=\alpha$, $f(b)=\beta$ и при $a \leqslant x \leqslant b$ функция $y=f(x)$ монотонно возрастает.

Докажите, что: а) $f(g(x))=x$, $g(f(x))=x$; б) функции $F(x)=f(f(x))$ и $G(x)=g(g(x))$ взаимно-обратны. Каковы области определения функций $f(g(x))$, $g(f(x))$; $F(x)$, $G(x)$?

§ 7. Преобразования графиков функций

Выше мы уже неоднократно встречались с «преобразованиями графиков», с переходами от графика функции к иной линии, родственной первоначальной. Здесь мы снова вернемся к этому вопросу с тем, чтобы рассмотреть его систематичнее.

Проще всего решить вопрос о сдвиге (параллельном переносе) кривой. Пусть мы имеем график функции $y=f(x)$; как будет выглядеть функция, график которой получается из нашего графика смещением на b единиц вверх? Ответ на этот вопрос почти очевиден: ясно, что если рассмотреть две кривые

$$y=f(x) \text{ и } y=f(x)+b,$$

то каждой точке $A(x_0, y_0)$ первого графика будет отвечать точка $A_1(x_0, y_1)$ второго графика, расположенная над точкой A, выше ее на b (рис. 1, число b мы здесь считаем положительным).

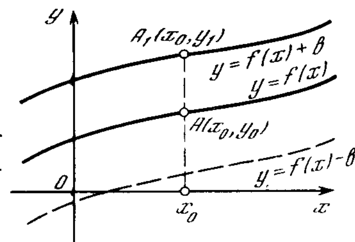


Рис. 1.7.1

По-другому это можно описать так: если наши две функции имеют вид

$$y=f(x) \text{ и } y-b=f(x),$$

то вторая расположена на b единиц выше первой, т. е. замена в уравнении линии величины y на $y-b$ равносильна сдвигу линии на b единиц вверх.

Может показаться, что мы напрасно повторили одно и то же утверждение двумя почти не отличающимися способами: поднять кривую — это совершить следующее преобразование ее уравнения: перейти от $y=f(x)$ к $y=f(x)+b$, или от $y=f(x)$ к $y-b=f(x)$. Однако вторая формулировка с заменой y на $y-b$ удобна для случая, когда кривая задана «неявным» образом, т. е. уравнением вида $F(x, y)=0$, не разрешенным относительно y. Так, например, уравнение окружности S радиуса 1 с центром в начале координат удобно записать так:

$$x^2+y^2=1, \text{ или}$$

$$F(x, y)=x^2+y^2-1=0 \tag{1.7.1}$$

(ср. с формулой (2.1) для расстояния OA, где $A=A(x, y)$). Заменив здесь y на $y-b$, мы придем к уравнению

$$x^2+(y-b)^2-1=0$$

единичной окружности с центром в точке $Q_1(0, b)$ — окружности, полученной из S сдвигом на b единиц вверх (рис. 2).

Аналогично (мы подробно обсудили это в § 4) замена в уравнении $F(x, y)$ кривой аргумента x на $x-a$ равносильна сдвигу кривой на a единиц вправо. Так, уравнение

$$(x-a)^2+y^2-1=0$$

определяет окружность с центром в точке $Q_2(a, 0)$ — окружность, получаемую из первоначальной окружности S сдвигом вправо на a (см. рис. 2).

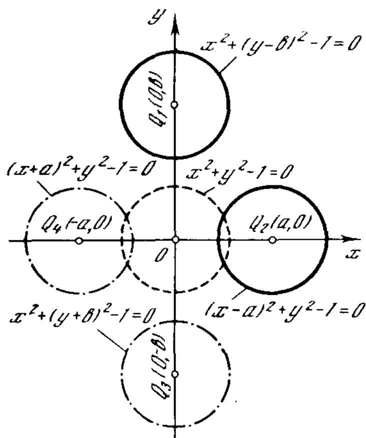


Рис. 1.7.2

Ясно, что величины a и b могут быть и отрицательны; так, например, переход от уравнения $y=f(x)$ к уравнению $y=f(x)-b$ (где по-прежнему $b>0$) или замена y в уравнении линии на $y+b$ означает, что мы опускаем график

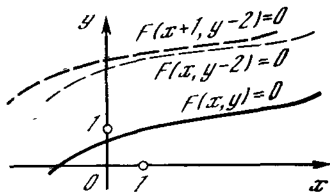


Рис. 1.7.3

линии на b единиц в н и з (см. линию, изображенную пунктиром на рис. 1). Аналогично этому линия $F(x+a, y)=0$ получается из линии $F(x, y)=0$ сдвигом на a единиц в л е в о. Так, центрами окружностей $x^2+(y+b)^2-1=0$ и $(x+a)^2+y^2-1=0$ являются точки $Q_3(0, -b)$ и $Q_4(-a, 0)$ (см. рис. 2). Указанные сдвиги можно и комбинировать: например, линия $F(x+1, y-2)=$ $=0$ получается из линии $F(x, y)=0$ сдвигом на 1 единицу влево и на 2 единицы вверх (рис. 3).

П р и м е р. Рассмотрим так называемую *дробно-линейную* функцию

$$y=\frac{ax+b}{cx+d}, \quad \text{где } c\neq 0. \tag{1.7.2}$$

Многие экспериментально наблюдаемые зависимости выражаются — точно или приближенно — функциями такого строения; поэтому важно знать, как можно упростить запись этой функции и как изобразить ее график.

Ясно, что если числитель и знаменатель стоящей в правой части (2) дроби пропорциональны, т. е. $a:c=b:d$, или $ad=bc$, то функция (2) постоянна — она имеет вид $y=k$, где $k=\frac{a}{c}\left(=\frac{b}{d}\right)$. Поэтому интерес представляет лишь случай, когда $ad\neq bc$. Мы утверждаем, что в этом случае графиком зависимости (2) является г и п е р б о л а, и, следовательно, запись (2) означает обратную пропорциональность двух величин.

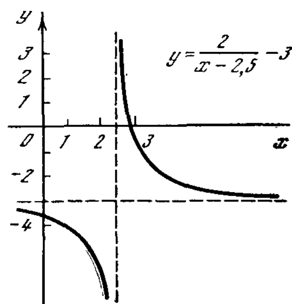


Рис. 1.7.4

Предоставив разбор общего случая читателю (см. упр. 6), мы ограничимся здесь частным случаем, когда

$$y=\frac{19-6x}{2x-5}. \tag{1.7.2а}$$

Выделим из стоящей в правой части дроби целую часть:

$$y=\frac{(15-6x)+4}{2x-5}=\frac{15-6x}{2x-5}+\frac{4}{2x-5}=$$
$$=-3+\frac{4}{2x-5}.$$

Последнее уравнение можно еще переписать так:

$$y+3=\frac{2}{x-5/2}, \tag{1.7.2б}$$

откуда видно, что линия (2а) (или (2б)) есть график обратной пропорциональной зависимости (с коэффициентом 2) величин $y+3$ и $x-5/2$ (рис. 4).

Выясним теперь, как надо изменить уравнение кривой для того, чтобы растянуть ее в c раз вдоль оси y, т. е. увеличить в c раз все ее вертикальные раз-

меры [1]. Очевидно, для этого необходимо заменить исходное уравнение $y=f(x)$ новым уравнением $y=cf(x)$ (мы будем писать: $y_0=f(x)$ и $y_1=cf(x)$, поскольку это две разные кривые). Тогда при том же x величина y_1 будет в c раз больше величины y_0, т. е. наша кривая растянется в вертикальном направлении в c раз.

В качестве примера обратимся снова к уравнениям прямых, проходящих через начало координат. Уравнение биссектрисы I и III координатных углов имеет вид $y_0=x$. Уравнение же

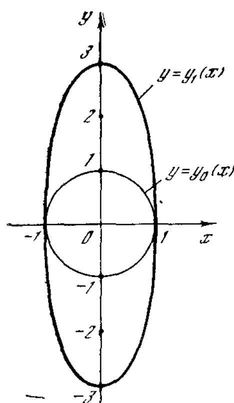


Рис. 1.7.5

$y_1=10x$ соответствует прямой, идущей более круто, у которой при заданном x ордината больше в 10 раз (см. рис. 3.4).

Переход от $y_0=f(x)$ к $y_1=cf(x)$ можно описать и так: в уравнении кривой $y_0=f(x)$ мы заменяем y_0 на $\frac{y_1}{c}$, т. е. пишем $\frac{y_1}{c}=f(x)$. Тогда зависимость y_1 (y «новое») от x характеризуется тем, что кривая вытянута в c раз по вертикали по сравнению с кривой $y_0(x)$ (где y_0 — это y «старое»).

Снова необходимость двух формулировок, тождественность которых очевидна (ведь ясно, что уравнения $y=cf(x)$ и $\frac{y}{c}=f(x)$ не отличаются одно от другого!), вызвана тем, что правило замены y на $\frac{y}{c}$ оказывается более удобным, если требуется «растянуть в c раз от оси x» кривую, уравнение $F(x, y)=0$ которой не разрешено относительно y. Достаточно подставить в новое уравнение $\frac{y}{c}$ вместо y — и мы придем к требуемому результату: $F\left(x, \frac{y}{c}\right)=0$.

Обратимся, например, снова к единичной окружности S с уравнением $x^2+y^2-1=0$. Как записать уравнение кривой, получаемой, если вытянуть S в 3 раза по вертикали (рис. 5; новая кривая обозначена через $y_1(x)$)? По правилу, которое мы только что сформулировали, для этого надо в уравнении окружности заменить y на $\frac{y}{3}$. При этом получим

$$x^2+\left(\frac{y_1}{3}\right)^2-1=0 \quad (1.7.3)$$

(мы пишем здесь y_1, а не y, чтобы отличить кривую (3) от окружности (1)).

Кривая, которая получается из окружности растяжением последней от горизонтального (или вертикального) ее диаметра (см., впрочем, упр. 5), называется *эллипсом* [2]; таким образом от окружности (1) (на рис. 5 эта окружность обозначена как $y_0(x)$) мы перешли к эллипсу (3).

В данном случае уравнения (1) и (3) легко решаются:

$$y_0=\sqrt{1-x^2} \qquad y_1=3\sqrt{1-x^2},$$

и наглядно видно, что $y_1=3y_0$ при одинаковых x. Но правило, утверждающее, что замена y на $\frac{y}{c}$ приводит к вытягиванию кривой в c раз по вертикали, справедливо и для кривых, заданных таким сложным уравнением $F(x, y)=0$, которое никак нельзя решить алгебраически относительно y, например уравнением

$$x+y\lg y=0. \quad (1.7.4)$$

Линия, получающаяся из линии (4) увеличением всех вертикальных размеров втрое, описывается следующим уравнением:

$$x+\frac{1}{3}y\lg\left(\frac{y}{3}\right)=0. \quad (1.7.4a)$$

Утверждение о смысле замены y на $\frac{y}{c}$ легко переносится и на координату x. При

[1] Рекомендуем читателю для простоты здесь и далее пока считать, что $c>1$; на случае $c<1$ (и даже $c<0$) мы еще специально остановимся ниже.

[2] Коэффициент c растяжения может быть и равным 1 (разумеется, такое «тождественное растяжение» кривую не меняет); в соответствии с этим окружность обычно считают частным случаем эллипса.

замене в уравнении кривой $F(x_0, y)=0$ аргумента x_0 на $\frac{x}{c}$, т. е при переходе к уравнению $F\left(\frac{x_1}{c}, y\right)=0$, линия растягивается вдоль оси x в c раз: при одинаковом y значение x_1 в c раз больше значения x_0.

Начнем не с доказательства, а с примеров:

$$y=x_0 \quad \text{и} \quad y=\frac{x_1}{10}=0{,}1x_1$$

(см. рис. 3.4). Первая прямая идет под углом 45° к оси x, вторая идет более полого.

Второй пример:

$$x_0^2+y^2-1=0, \quad \left(\frac{x_1}{2}\right)^2+y^2-1=0.$$

Первое уравнение описывает окружность S радиуса 1, вторая кривая получается из первой растяжением вдоль оси x в 2 раза (рис. 6). Ясно, что и здесь новая кривая представляет собой э л л и п с.

Доказательство нашего утверждения почти очевидно. Если решить уравнения первой и второй кривых

$$y=f(x_0) \quad \text{и} \quad y=f\left(\frac{x_1}{c}\right), \qquad \textbf{(1.7.5)}$$

относительно x, то получим

$$x_0=\varphi(y) \quad \text{и} \quad \frac{x_1}{c}=\varphi(y),$$

т. е. $x_1=c\varphi(y)=cx_0$, **(1.7.5а)**

где φ — функция, обратная функции f (см. § 6). Это соответствует исходной формулировке: при замене x на $\frac{x}{c}$ кривая растягивается вдоль оси x в c раз. При этом важно, что в (5) f представляет собой одну и ту же функцию в формулах с x_0 и с x_1; поэтому и в (5а) φ в формулах с x_0 и с x_1 — это одна и та же функция.

Вот еще один пример: пусть

$$y=10^{x_0} \quad \text{и} \quad y=10^{x_1/2}. \qquad \textbf{(1.7.6)}$$

Но степенной функции обратна логарифмическая функция, поэтому здесь

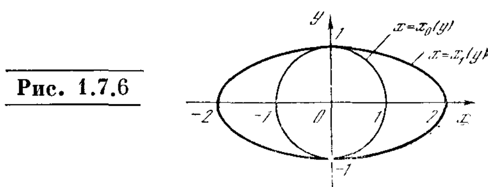
$$x_0=\lg y \quad \text{и} \quad \frac{x_1}{2}=\lg y,$$

т. е. $x_1=2\lg y$. **(1.7.6а)**

Таким образом, график функции $y=10^{x/2}$ получается из графика функции $y=10^x$ растяжением в 2 раза в направлении оси абсцисс.

А что означает замена x на kx в уравнении $y=f(x)$? Для того чтобы воспользоваться сформулированным выше правилом, вспомним, что умножение на k — это то же самое, что деление на $1/k$: ведь $kx=\frac{x}{1/k}$. Значит, $1/k$ играет роль величины c в предыдущих формулах, и если, например $k=1/2$, то $c=1/k=2$. Значит, замена x_0 на $0{,}5x_1$ есть замена x_0 на $x_1/2$: она приводит к растяжению кривой вдоль оси x в 2 раза. Если же $k=3$, то $c=1/k=1/3$.

Рис. 1.7.6



Таким образом, замена x на $3x$ — это то же, что и замена x на $\frac{x}{1/3}$.

Что означают эти подстановки геометрически? В случае, когда $c>1$, результат формулируется так: при замене $y\to\frac{y}{c}$ в уравнении кривой линия в c раз растягивается по высоте; при замене же $x\to\frac{x}{c}$ линия растягивается в c раз в горизонтальном направлении.

В случае, когда $0<c<1$, в наших рассуждениях, по существу, не меняется ничего, однако слова «увеличить» и «растянуть» здесь уже не подходят: ведь «увеличить в 2 раза» какую-то величину — это значит заменить ее на вдвое большую, а вот «увеличить» или «растянуть» в 1/3 раза — это значит умножить на 1/3, т. е. не увеличить, а в 3 раза уменьшить, в 3 раза сжать. Таким образом, здесь снова при замене $y\to\frac{y}{c}$ вертикальные размеры изменяются в c раз, что в данном случае представляет собой с ж а т и е кривой: так, при $c=0{,}5$ переход от линии $y=f(x)$ к линии $y=cf(x)=0{,}5f(x)$ означает уменьшение ее высоты вдвое. То же относится и к подстановке $x\to\frac{x}{c}$: , при $0<c<1$ эта подстановка равносильна с ж а т и ю кривой вдоль оси x.

Выясним теперь смысл подстановок $y \to \frac{y}{c}$ $\left(\text{или } x \to \frac{x}{c}\right)$ при отрицательном c. Такую замену можно провести в два приема. Обозначим $c=-b=(-1)b$, где уже $b>0$, и сделаем две последовательные замены:

$$y_0 \to \frac{y_1}{b} \to$$

$$\to \frac{y_2}{(-1)b}\left(=-\frac{y_2}{b}, \quad \text{так что} \quad y_2=-y_1\right).$$

Первая операция — замена y_0 на y_1/b, где $b>0$, уже разбиралась выше, она

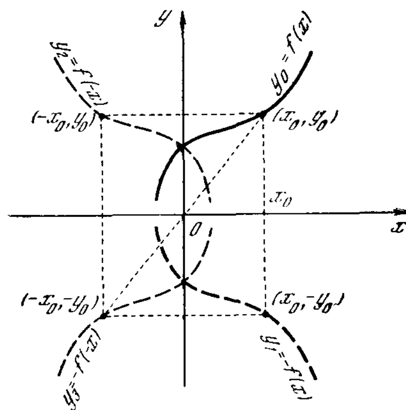


Рис. 1.7.7

приводит к изменению вертикальных размеров в b раз. Остается выяснить, какой смысл имеет изменение знака y, т. е. замена y на $-y$. Но очевидно, что соответствующие друг другу точки линий $y_0=f(x)$ и $y_1=-f(x)$ симметричны относительно оси x (рис. 7). Поэтому линия $y_1=-f(x)$ получается из линии $y_0=f(x)$ отражением от оси x, и то же можно сказать о линиях $F(x, y_0)=0$ и $F(x, -y_1)=0$, где второе уравнение получается из первого изменением знака y.

Аналогично этому, поскольку точки (x, y) и $(-x, y)$ симметричны относительно оси y, замена x на $-x$ в уравнении кривой (переход от линии $F(x, y_0)=0$ к линии $F(-x, y_2)=0$) заменяет первоначальную линию симметричной ей относительно оси y. Наконец, переход от уравнения $F(x, y_0)=0$ к уравнению $F(-x, -y_3)=0$ сводится к последовательным изменениям знака y (симметрии относительно оси x) и к изменению знака x (симметрия относительно оси y). Но последовательное осуществление двух отражений — от оси x и от оси y — равносильно отражению линии от начала координат O (0, 0) (см. рис. 7). Таким образом, одновременная замена в уравнении линии x на $-x$ и y на $-y$ равносильна симметрии относительно точки O.

Как ни просты, «арифметичны» все эти соображения, начинающие (которым адресована эта книга) в них часто ошибаются.

Рассмотрим пример:

$$F(x, y)=(x-3)^2+(y-5)^2-4=0.$$

Это уравнение окружности радиуса 2 с центром в точке с координатами $x=3$,

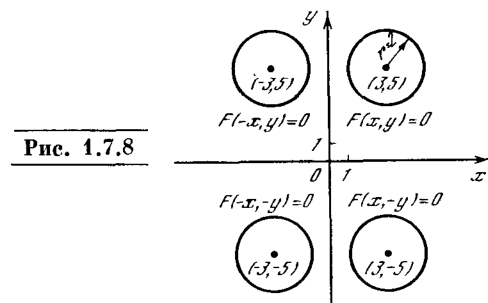


Рис. 1.7.8

$y=5$. На рис. 8 изображены также линии

$$F(x, -y)=(x-3)^2+(-y-5)^2-4=0,$$
$$F(-x, y)=(-x-3)^2+(y-5)^2-4=0,$$
$$F(-x, -y)=(-x-3)^2+(-y-5)^2-4=0.$$

Здесь буква F во всех случаях обозначает одну и ту же функцию двух пере-

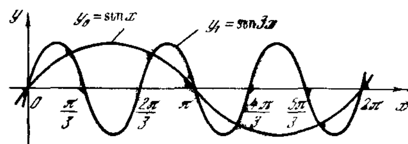


Рис. 1.7.9

менных. Проследите, как трансформируется (изменяется) кривая (окружность) при замене x на $-x$, или y на $-y$, или при одновременной замене $x \to -x$, $y \to -y$. Ясное понимание изложенных правил позволит Вам, построив одну какую-то кривую $y=f(x)$ или $F(x, y)=0$, представить себе, как выглядят графики всех функций

$$\frac{y-b}{d}=f\left(\frac{x-a}{c}\right) \text{ или } F\left(\frac{x-a}{c}, \frac{y-b}{d}\right)=0$$

с любыми значениями постоянных a, b, c, d.

Вот еще два простых, но принципиальных примера. На рис. 9 изображены две кривые:

$$y = \sin x \quad \text{и} \quad y = \sin 3x, \tag{1.7.7}$$

где, разумеется, x — радианная мера угла. У второй кривой все горизонтальные размеры в 3 раза меньше, чем у первой.

Зависимость $y=\sin x$ *периодическая*: при $x=2\pi\approx 6{,}3$ (радиан, что соответствует углу 360° в градусном измерении) синус имеет такое же значение, как и при $x=0$ (добавление 2π к любому углу не меняет значения синуса). Таким образом, график первой из кривых (7) переходит в себя при сдвиге вдоль оси x (в любом направлении) на величину 2π — это и означает, что функция $\sin x$ периодическая, с периодом 2π. Функция $y=\sin 3x$ тоже периодическая, но у нее период меньше в 3 раза; достаточно, чтобы x изменилось на $\frac{2\pi}{3}\approx 2{,}1$ радиана, тогда $3x$ (тот угол, синус которого мы откладываем по оси ординат) меняется на 2π и $\sin 3x$ возвращается к прежнему значению:

$$\sin 3x = \sin 3\left(x + \frac{2\pi}{3}\right).$$

Продумайте на этом примере общее утверждение о том, что для периодической функции при замене x на kx в k раз уменьшается *период* и в k раз увеличивается *частота* функции, т. е. число периодов, приходящихся на единицу длины.

Второй пример связан с логарифмическими (и показательными) функциями. Рассмотрим две кривые

$$y_1 = \log_a x \quad \text{и} \quad y_2 = \log_b x, \tag{1.7.8}$$

или, что то же самое,

$$x = a^{y_1} \quad \text{и} \quad x = b^{y_2}, \tag{1.7.8a}$$

где мы считаем, что $a,\ b > 1$. Аналогично анализу зависимостей (6), (6а) устанавливается (см. § 4.9), что вторая кривая (8) получается из первой кривой растяжением вдоль оси y в $c=\log_b a = \frac{1}{\log_a b}$ раз, где число c называется *модулем перехода* от логарифмов при основании a к логарифмам при основании b. Аналогично любые две показательные функции

$$y = a^x \quad \text{и} \quad y = b^x$$

связаны между собой подобным же образом: вторая из них получается из первой растяжением в $c=\log_b a$ раз в направлении оси x.

Со сказанным выше связаны еще два простых, но важных понятия. Кривая, уравнение $F(x,\ y)=0$ которой сохраняет свой вид при замене $x \to -x$ (скажем, парабола $y=ax^2$), симметрична относительно оси y. Кривая, уравнение которой сохраняет свой вид при замене $x \to -x$ и $y \to -y$ (скажем,

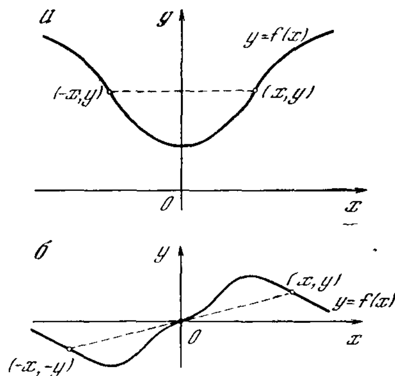


Рис. 1.7.10

гипербола $y=\frac{1}{x}$ или кубическая парабола $y=x^3$), симметрична относительно начала координат. Наконец, кривая, уравнение $F(x,\ y)=0$ которой переходит в себя при подстановке $y \to -y$ (например, полукубическая парабола $y^2=x^3$ (см. рис. 5.7), симметрична относительно оси x. Все эти утверждения автоматически следуют из геометрического смысла подстановок $x \to -x$; $x \to -x$ и $y \to -y$; $y \to -y$.

Функция $y=f(x)$, график которой симметричен относительно оси y (рис. 10, *а*), называется *четной*; если же график функции симметричен относительно точки O (рис. 10, *б*), то функция называется *нечетной*. Эти названия связаны с тем, что параболы четного порядка $y=ax^{2m}$ являются четными функциями, а параболы нечетного порядка $y=x^{2m+1}$ — нечетными (ср. § 5).

Наконец, остановимся еще на одном очень важном обстоятельстве. Необходимо понимать (соображение, на которое обычно, к сожалению, не обращают должного внимания), что в прикладных задачах процедура изменения в постоянном отношении всех ординат y или всех абсцисс x связана (всего лишь!)

с изменением единиц измерения величин y или x, так что полученные один из другого таким путем графики выражают одну и ту же зависимость между y и x, так сказать, «с точностью до изменения масштабов» (т. е. системы единиц [3]).

Пусть, например, по оси абсцисс Ox откладывается время, а по оси ординат Oy — расстояние (рис. 11); зависимость $y=f(x)$ (или $z=f(t)$, как мы чаще будем писать) задает закон движения. Однако при этом y и x измеряются в разных единицах, так что никакое сравнение

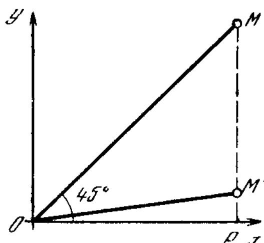


Рис. 1.7.11

их невозможно: ведь нельзя же приравнять между собой 1 см и 1 с или сказать, какая из этих величин больше, а какая меньше. Утверждение, например, о том, что для данной точки $M=M(x, y)$ отрезок OM составляет 45° с осью абсцисс (рис. 11), является бессодержательным: изменив, скажем, единицы расстояний (перейдя от сантиметров к метрам), мы заменим точку M на новую точку M', отвечающую тому же месту на прямой и тому же моменту времени; однако теперь равенство $\widehat{xOM}=45°$ уже не будет иметь места, а ведь для нас точка M' полностью равноправна M. Таким образом, в физических рассмотрениях изображенные на рис. 5 или 6 эллипсы ничем не отличаются от изображенной на тех же рисунках окружности, — и если эта линия нас особенно интересует, то единицы измерения, пожалуй, следует выбрать так, чтобы она оказалась именно окружностью.

Аналогично этому любую параболу $y=ax^2+bx+c$ прежде всего можно, как мы видели выше, сдвигом (т. е. переходом от координат x и y к новым координатам $x_1=x+p$ и $y_1=y+q$, где p и q подобраны подходящим образом) свести к параболе $y_1=ax_1^2$. Но ведь подобный сдвиг физически означает не что иное, как изменение начала отсчета единиц, скажем начального момента времени (момента $t=0$) и начала отсчета расстояний (начального пункта $z=0$), что не затрагивает интересующий нас физический процесс. После этого изменением масштаба по оси y_1 (выбором новой единицы измерения ординат) мы можем свести параболу $y_1=ax_1^2$ к «единичной параболе» $Y=\pm X^2$ (или даже именно к параболе $Y=X^2$, если мы имеем право располагать выбором «положительного направления» вдоль оси y [4]); именно этот смысл имело сделанное выше (см. сноску 3 в § 4) замечание о том, что параболу $y=x^2$ можно считать «общей».

Точно так же любую дробно-линейную зависимость $y=\frac{kx+l}{mx+n}$ можно изменением начала отсчета абсцисс x и ординат y и изменением масштаба по оси y свести к единичной гиперболе $Y=\frac{1}{X}$ (см. упр. 6); любую кубическую зависимость $y=ax^3+bx^2+cx+d$ можно свести к одной из трех зависимостей $y=x^3+x$, $y=x^3-x$ и $y=x^3$, далее уже друг к другу не сводимых, и т. д.

Остановимся на доказательстве более сложного утверждения, связанного с кубическими зависимостями

$$y=ax^3+bx^2+cx+d=$$
$$=a(x^3+px^2+qx+r), \tag{1.7.9}$$

где $p=\frac{b}{a}$, $q=\frac{c}{a}$, $r=\frac{d}{a}$ (здесь, разумеется, $a\neq 0$). В § 5 мы видели, что подстановка $X=x+\frac{p}{3}$ переводит зависимость (9) в зависимость

$$y=a(X^3+PX+Q),\quad \text{или}\quad Y=aX^3+BX=$$
$$=a(X^3+kX), \tag{1.7.9a}$$

где $Y=y-aQ$, $B=aP$ и $k=B/a\,(=P)$, а $P=q-\frac{p^2}{3}$, $Q=\frac{2}{27}p^3+r-\frac{1}{3}pq$ (ср. формулы (5.6) и (5.6а)). При $k=0$ (т. е. если

[3] Впрочем, если величины x и y измеряются в одних единицах, то подстановка $Y=cy$ и физически значима: ведь здесь зависимости $y=f(x)$ и $y=cf(x)$ имеют разный смысл.

[4] В то время как для времени t мы имеем «естественное» направление роста, что отвечает принципиальной неравноправности прошлого и будущего, для расстояний z дело обстоит вовсе не так — и тут мы свободны в выборе направления возрастания z.

$B=0$) зависимость (9а) имеет вид $Y=aX^3$, или

$$y_1=x_1^3, \quad (1.7.10а)$$

где $x_1=X$ (координата X не меняется) и $y_1=\frac{Y}{a}$. Если $k>0$ (т. е. B и a одного знака), положим $x_1=\frac{X}{\sqrt{k}}$, т. е. $X=\sqrt{k}\,x_1$; тогда (9а) обратится в

$$Y=a\left(k^{3/2}x_1^3+k^{3/2}x_1\right)=ak^{3/2}(x_1^3+x_1),$$

т. е. здесь мы имеем закон

$$y_1=x_1^3+x_1, \quad (1.7.10б)$$

где $y_1=\frac{Y}{ak\sqrt{k}}$. Наконец, при $k<0$ (в случае, когда a и B разных знаков) аналогичная использованной выше подстановка $x_1=\frac{X}{\sqrt{|k|}}$, т. е. $X=\sqrt{|k|}\,x_1$, переводит равенство (9а) в

$$Y=a\left(|k^{3/2}|\,x_1^3+k\,|k|^{1/2}x_1\right),$$

т. е.

$$y_1=x_1^3-x_1, \quad (1.7.10в)$$

где $y_1=\frac{-Y}{ak\sqrt{|k|}}$. Таким образом, для физика любая естественнонаучная зависимость (9) между (разноразмерными!) величинами y и x равносильна одному из трех законов (10а)—(10в) (см. рис. 5.1, *а*, 5.2 и 5.3; ср. упр. 8 и 9).

Поясним сказанное на конкретном примере. Хорошо известно, что зависимость от времени t высоты h камня, брошенного вверх с башни высоты w с начальной скоростью v, выражается формулой

$$h=w+vt-\frac{g}{2}t^2, \quad (1.7.11)$$

где $g\approx 9{,}8$ м/с² — ускорение свободного падения. Если, однако, принять за начало отсчета высот не уровень Земли, а самую большую достигнутую камнем высоту h_0, а за начало отсчета времени принять момент t_0 достижения камнем этой высоты, то в новых координатах $h_1=h-h_0$, $t_1=t-t_0$ уравнение (11) движения камня заметно упростится:

$$h_1=-\frac{gt_1^2}{2}. \quad (1.7.11а)$$

Но и коэффициент $-\frac{g}{2}$ в уравнении (11а) связан вовсе не с существом наблюдаемого процесса, а всего лишь с выбором системы единиц! Прежде всего мы можем упростить (11а), отбросив в правой части этой формулы знак «минус»; для этого достаточно условиться измерять расстояния h от точки $h=h_0$ не вверх, а вниз. Далее можно перейти к «натуральной системе единиц», в которой ускорение силы тяжести равно 2 (скажем, измерять время по-прежнему в секундах, а за единицу расстояния принять величину, численно равную $\frac{g}{2}$, т. е. $\approx 4{,}9$ м). Тогда в новых координатах $H=-\frac{2}{g}(h-h_0)$, $T=t_1=t-t_0$ закон (11) (или (11а)) примет «канонический» вид:

$$H=T^2.$$

Подобным образом встречающуюся в естественнонаучных законах логарифмическую или показательную зависимость мы всегда можем переписать так, чтобы основание системы логарифмов или основание степени было таким, как нам удобно, — ведь переход от одного основания к другому сводится лишь к замене единиц измерения фигурирующих в нашей зависимости величин (ср. со сказанным выше о логарифмических и показательных кривых). Поэтому каждый раз, когда мы встречаемся с логарифмической зависимостью $y=\log_c x$ двух величин, мы свободно можем считать логарифмы десятичными, или натуральными (логарифмы по основанию $e\approx 2{,}272$; см. § 4.9), или, например, двоичными [5]; при рассмотрении показательных зависимостей $y=d^x$ обычно считают основание степени d равным e и т. д.

Упражнения

1.7.1. Постройте кривые $\frac{x^2}{4}+\frac{y^2}{9}-1=0$; $\frac{(x+3)^2}{4}+\frac{(y-5)^2}{9}-1=0$; $\frac{(x-3)^2}{9}+\frac{(y+5)^2}{4}-1=0$ (при этом естественно воспользоваться тем, что линия $x^2+y^2-1=0$ представляет собой единичную окружность).

[5] Так, например, в *теории информации*, где количество информации выражается логарифмическими функциями переменных, принято считать все логарифмы *двоичными*, что, однако, не имеет никакого принципиального смысла, а связано лишь с принятой системой измерения величин (с измерением количеств информации в так называемых *битах*). См. по этому поводу, например, книгу: *Яглом А. М., Яглом И. М.* Вероятность и информация. М.: Наука, 1973.

1.7.2. Постройте график функции $y=\sin x$, взяв, например, восемь значений x в интервале от $-\pi$ до $+\pi$, различающихся между собой на целые кратные величины $0{,}25\pi$. Воспользовавшись этим графиком, постройте кривые: а) $y=2\sin x$; б) $y=\sin 0{,}5x$; в) $y=3\sin 3x$; г) $y=\cos x$; д) $y=\cos x+\sin x$ $\left(=\sqrt{2}\sin\left(x+\frac{\pi}{4}\right)\right)$; е) $y=\cos^2 x$ $(=1/2+{}^1/_2\cos 2x)$; ж) $y=\sin^2 x$ $(=1/2-{}^1/_2\cos 2x)$. [Указание. Все эти линии можно построить, сдвигая, растягивая или сжимая синусоиду $y=\sin x$; для решения упр. г), е), ж) воспользуйтесь тождеством $\cos x=\sin\left(x+\frac{\pi}{2}\right)$.]

1.7.3. Постройте кривые: а) $y=\pm\sqrt{x^2-1}$; б) $y=\pm\sqrt{x^2+1}$; в) $y=2\pm\sqrt{(x-1)^2-1}$; г) $4y^2+4y-x^2=0$. [Указание. Кривую а) постройте по точкам, выбрав в интервале от -5 до $+5$ двадцать значений x, различающихся на 0,5 (и кратных 0,5). Преобразовав уравнения а) и б) к виду $y^2-x^2+1=0$, соответственно $x^2-y^2+1=0$, заметим, что б) получается из а) перестановкой x и y. Для решения г) запишите уравнение этой кривой в виде $4\left(y+\frac{1}{2}\right)^2-x^2-1=0$ и получите ее, сдвигая и сжимая кривую б).]

1.7.4. Выпишите уравнение кривой, полученной из кривой $y=x^3-x$ (см. рис. 5.3): а) сжатием в направлении оси Oy с коэффициентом 1/3 (т. е. уменьшением в 3 раза всех вертикальных размеров); б) подстановкой $x\to\frac{x}{2}$; в) подстановкой $y\to-\frac{y}{2}$. Изобразите полученные кривые на чертеже.

1.7.5. Докажите, что растяжение (или сжатие) единичной окружности $x^2+y^2-1=0$ к любой прямой с каким угодно коэффициентом растяжения (сжатия) переводит ее в эллипс. [Указание. Эту задачу можно сформулировать так: растяжение или сжатие любой окружности $(x-a)^2+(y-b)^2-r^2=0$ к оси x, т. е. подстановка $y\to ky$, переводит эту окружность в эллипс.]

1.7.6. Докажите что график произвольной дробно-линейной зависимости $y=\dfrac{ax+b}{cx+d}$ между переменными y и x, где $c\neq 0$, $\Delta=ad-bc\neq 0$, может быть получен из графика обратной пропорциональности $y=\frac{k}{x}$, где $k=-\frac{\Delta}{c^2}$, переносом последнего графика на $\frac{d}{c}$ влево $\left(\text{т. е. при } \frac{d}{c}<0 \text{ переносом на } \left|\frac{d}{c}\right| \text{ вправо}\right)$ и на $\frac{a}{c}$ вверх.

1.7.7. Докажите, что каждую функцию можно представить в виде суммы четной и нечетной функций. [Указание. Воспользуйтесь тождеством $f(x)=\frac{1}{2}(f(x)+f(-x))+\frac{1}{2}(f(x)-f(-x))$.]

1.7.8. Докажите, что никакие две из кривых (10а)—(10в) несводимы одна к другой подстановками вида $x'=ax+p$, $y'=bx+q$ (где, разумеется, $a\neq 0$, $b\neq 0$), равносильными изменению начала отсчета величин x и y (переносу начала координат), сопровождаемому изменением единиц измерения x и y.

1.7.9. Укажите правило, позволяющее судить по коэффициентам a, b, c, d зависимости (9) о том, сводится ли она к виду (10а), (10б) или (10в).

§ 8. Параметрическое задание линий

Рассмотрим процесс движения материальной точки $M=M(x, y)$. Каждая из координат x, y точки будет здесь меняться с течением времени t: движение M зададут две функции $x(t)$ и $y(t)$ времени, например

$$x=\cos t,\quad y=\sin t. \tag{1.8.1}$$

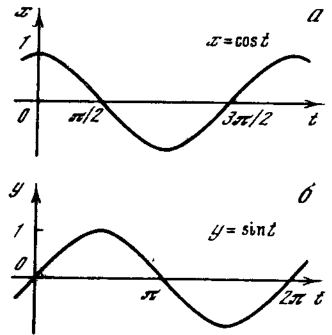


Рис. 1.8.1

Эти зависимости можно изобразить графически в виде двух кривых, откладывая на одном чертеже по оси абсцисс t, а по оси ординат x, а на другом чертеже — по оси абсцисс t, а по оси ординат y (рис. 1).

Зададимся теперь вопросом о траектории точки M. Каждому значению t отвечают свои значения $x(t)$ и $y(t)$ координат; спрашивается, какую кривую опишет точка $M=M(x(t), y(t))=$

$=M(t)$ при изменении t. Для ответа на этот вопрос можно из двух уравнений $x=x(t)$, $y=y(t)$ исключить величину t; тогда мы получим выражение, в которое будут входить только y и x, т. е. уравнение вида $y=y(x)$ или $F(x, y)=0$. После этого будем строить кривую как обычно, задаваясь различными x и находя соответствующие y.

Так, в приведенном примере найдем:

$$x^2+y^2=\cos^2 t+\sin^2 t=1,$$

т. е.

$$y=\pm\sqrt{1-x^2}, \quad \text{или} \quad x^2+y^2-1=0,$$

так что кривая (1) в плоскости x, y представляет собой о к р у ж н о с т ь (рис. 2, *а*).

Однако часто даже сравнительно простые выражения для $x(t)$ и $y(t)$ при попытке исключения из них t приводят к столь сложным формулам, что никто не решится идти этим путем. Так, например, если

$$x=x(t)=a_1t^4+b_1t^3+c_1t^2+d_1t+e_1,$$

$$y=y(t)=a_2t^4+b_2t^3+c_2t^2+d_2t+e_2,$$

то для исключения t нужно решить уравнение четвертой степени; это приводит к выражениям, выписывать которые в силу их громоздкости просто невозможно. Однако построить кривую в плоскости x, y можно и не исключая t: достаточно задаться разными значениями t и найти значения x и y, отвечающие каждому из них. Вот, например, как выглядит таблица значений функций $x(t)$ и $y(t)$, отвечающая простому примеру (1):

t	0	$\pi/4$	$\pi/2$	$3\pi/4$	π
$x=\cos t$	1	0,7	0	−0,7	−1
$y=\sin t$	0	0,7	1	0,7	0

t	$5\pi/4$	$3\pi/2$	$7\pi/4$	2π
$x=\cos t$	−0,7	0	0,7	1
$y=\sin t$	−0,7	−1	−0,7	0

Очевидно, брать t больше 2π нет надобности: далее значения x и y будут повторяться. Теперь с помощью этой таблицы строим точки кривой. При этом мы используем только значения x и y. Те значения t, при которых эти x и y вычислены, при построении точек не используются — «мавр сделал свое дело, мавр может уйти».

Такой способ задания кривой, или, что то же самое, функциональной зависимости $y(x)$, двумя функциями $x(t)$ и $y(t)$ называется *параметрическим*, а величина t называется *параметром* [1]. В физических задачах параметр часто имеет вполне определенный смысл. Так, в том примере, с которого мы начали этот параграф, t может иметь смысл времени; в этом слу-

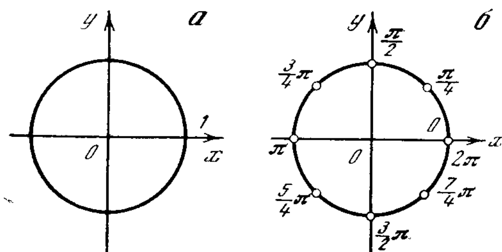


Рис. 1.8.2

чае $x(t)$ и $y(t)$ суть две функции времени. При этом можно интересоваться исключительно формой той кривой, которую опишет точка $M(x, y)$, но можно интересоваться и тем, с какой скоростью движется эта точка и какое положение будет она занимать в разные моменты времени. Для того чтобы узнать это, нам необходимо сохранить значения параметра и при каждой точке $M=M(t)$ выписать число t (разумеется, практически мы можем выписать на чертеже лишь конечное число значений; см., например, рис. 2, *б*). Таким образом, рисунок «параметризованной кривой» доставляет нам больше информации, чем одна только траектория точки M (ср. рис. 2, *а* и рис. 2, *б*).

Рассмотрим для примера следующие кривые:

1) $x=\cos t,\quad y=\sin t;$

2) $x=\cos t,\quad y=-\sin t;$

3) $x=\cos 3t,\quad y=\sin 3t;$

4) $x=\sin 3t,\quad y=\cos 3t.$

[1] Слово «параметр» (по-русски его смысл можно передать, например, выражением «мерило») было специально для этой цели придумано древнегреческими математиками; точного перевода оно не имеет.

Во всех случаях, исключая t, получим одну и ту же кривую — окружность единичного радиуса в плоскости x, y, уравнение которой имеет вид $x^2+y^2=1$.

В чем же различие этих четырех случаев, как в каждом из них происходит движение? В первых трех случаях в момент $t=0$ точка находится на оси абсцисс, в последнем случае — на оси ординат. В случаях 1) и 3) точка (можно представлять здесь себе материальную частицу) движется по кругу против часовой стрелки, а в случаях 2) и 4) — по часовой стрелке. По абсолютной величине угловая скорость вращения (а значит, и линейная скорость $|v|$, ибо радиус круга равен 1) в случаях 1) и 2) равна 1 радиан/с (т. е. $|v|=1$ м/с, если радиус круга 1 м); в случаях же 3) и 4) имеем $|v|=3$ м/с (угловая скорость равна 3 радиан/с).

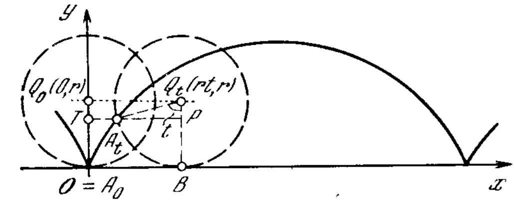


Рис. 1.8.3

Приведем еще один пример: построим кривую, по которой движется точка обода велосипедного колеса при движении велосипедиста в одном направлении — точка A окружности диска радиуса r при качении диска по прямой. Эта кривая называется *циклоидой* [2].

Примем горизонтальную прямую, по которой катится наш диск, за ось x; условимся еще считать, что в момент начала движения центр $Q=Q_0$ диска находится на оси y, имея координаты $(0,\ r)$. Будем считать скорость качения диска постоянной; например, положим, что в единицу времени диск поворачивается на единичный угол; в таком случае за время t диск смещается в положительном направлении оси x на расстояние rt (заметьте, что диск катится без скольжения, так что, повернувшись на угол t, он пройдет путь, равный дуге BA_t на рис. 3, т. е. путь rt). В этот момент центр Q_0 диска займет положение $Q_t\,(rt,\ r)$, а точка A_0 — положение A_t; нам требуется найти координаты точки A_t.

Из прямоугольного треугольника Q_tA_tP (см. рис. 3), в котором $A_t\widehat{Q_t}P=t$ и $Q_tA_t=r$, имеем:

$$A_tP=r\sin t,\quad Q_tP=r\cos t.$$

Отсюда, поскольку $BO=PT=rt$, $Q_tB=Q_0O=r$, получаем:

$$x=TA_t=TP-A_tP=rt-r\sin t,$$
$$y=OT=BP=Q_tB-Q_tP=r-r\cos t.$$

Окончательно,

$$x=r\,(t-\sin t),\quad y=r\,(1-\cos t)\quad (1.8.2)$$

(жирная линия на рис. 3 — это и есть циклоида).

Упражнения

1.8.1. Постройте кривую, заданную уравнениями: а) $x=\cos t$, $y=\sin 2t$; б) $x=\cos t$, $y=\sin 3t$. [Указание. Так как $\sin 3t$ меняется быстро, то нужно брать достаточно близкие значения t, например $t=0;\ 0{,}1;\ 0{,}2$, или, если у нас нет таблицы тригонометрических функций, аргумент которых измеряется в радианах, то $t=0;\ 5°;\ 10°;\ 15°;\ \ldots$]

1.8.2. Постройте кривую $x=\cos(5t+1)$, $y=\sin(5t+1)$.

1.8.3. Постройте кривую $x=\cos t$, $y=\cos\left(t+\frac{\pi}{4}\right)$.

1.8.4. Постройте кривую $x=\cos t$, $y=\cos t$.

1.8.5. Выразите параметрическую зависимость (2) явно — в виде функции $x=\varphi\,(y)$.

1.8.6. Окружность s радиуса r катится (без скольжения) по окружности S радиуса R; выпишите (параметрические) уравнения кривой, описываемой точкой A движущейся окружности s, если s расположена: а) вне S; б) внутри S или (при $R<r$), если S находится внутри s.

[Линия упр. а) называется *эпициклоидой*, а линия упр. б) — *гипоциклоидой.*]

Рассмотрите отдельно случаи: в упр. а) $r=R$; $r=\frac{1}{2}R$; $r=2R$; в упр. б) $r=\frac{1}{2}R$; $r=\frac{1}{3}R$; $r=\frac{1}{4}R$. (При $r=R$ эпициклоида называется *кардиоидой* (от греческого слова cardie — сердце); при $r=\frac{1}{3}R$ и $r=\frac{1}{4}R$ гипоциклоида называется соответственно *кривой Штейнера*[3] и *астроидой* (от латинского слова astrum — звезда). А что представляет собой гипоциклоида при $r=\frac{1}{2}R$?)

[2] От греческого слова kykloeides — кругообразный, порожденный кругом.

[3] Якоб Штейнер (1826—1863) — выдающийся немецкий геометр.

§ 9*. Некоторые дополнительные сведения из аналитической геометрии

Выше (в § 2) мы уже рассмотрели ряд геометрических задач, решаемых с применением координат, т. е. методами *аналитической геометрии*. Продолжим здесь эти рассмотрения.

Пусть $A_1(x_1,y_1)$, $A_2(x_2,y_2)$ и $A_3(x_3,y_3)$ — три точки на координатной плоскости (рис. 1). Как найти угол $A_2\widehat{A_1}A_3=\alpha$? В § 2 мы видели, что

$$\operatorname{tg}\alpha_{12}=\frac{y_2-y_1}{x_2-x_1} \quad \text{и} \quad \operatorname{tg}\alpha_{13}=\frac{y_3-y_1}{x_3-x_1},$$

где α_{12} и α_{13} — углы, образуемые отрезками A_1A_2 и A_1A_3 с осью x (см. формулу (2.5)). Но из рис. 1 с очевидностью вытекает, что

$$\alpha=A_2\widehat{A_1}A_3=\alpha_{13}-\alpha_{12}.$$

А так как по известной формуле тригонометрии

$$\operatorname{tg}(\beta-\gamma)=\frac{\operatorname{tg}\beta-\operatorname{tg}\gamma}{1+\operatorname{tg}\beta\operatorname{tg}\gamma},$$

то

$$\operatorname{tg}\alpha=\operatorname{tg}(\alpha_{13}-\alpha_{12})=\frac{\operatorname{tg}\alpha_{13}-\operatorname{tg}\alpha_{12}}{1+\operatorname{tg}\alpha_{13}\operatorname{tg}\alpha_{12}}=$$

$$=\frac{\dfrac{y_3-y_1}{x_3-x_1}-\dfrac{y_2-y_1}{x_2-x_1}}{1+\dfrac{y_3-y_1}{x_3-x_1}\dfrac{y_2-y_1}{x_2-x_1}},$$

или окончательно

$$\operatorname{tg}\alpha=\frac{(y_3-y_1)(x_2-x_1)-(y_2-y_1)(x_3-x_1)}{(x_3-x_1)(x_2-x_1)+(y_3-y_1)(y_2-y_1)}. \tag{1.9.1}$$

Таким образом, по координатам точек A_1, A_2 и A_3 можно определить угол $\alpha=A_2\widehat{A_1}A_3$.

Из формулы (1) с учетом известных соотношений $\sin\alpha=\sqrt{\dfrac{\operatorname{tg}^2\alpha}{1+\operatorname{tg}^2\alpha}}$ и $\cos\alpha=\sqrt{\dfrac{1}{1+\operatorname{tg}^2\alpha}}$ можно получить также формулы

$$\sin\alpha=$$
$$=\frac{(x_2-x_1)(y_3-y_1)-(x_3-x_1)(y_2-y_1)}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}\sqrt{(x_3-x_1)^2+(y_3-y_1)^2}}, \tag{1.9.2а}$$

$$\cos\alpha=$$
$$=\frac{(x_2-x_1)(x_3-x_1)+(y_2-y_1)(y_3-y_1)}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}\sqrt{(x_3-x_1)^2+(y_3-y_1)^2}}. \tag{1.9.2б}$$

Сравнительно сложная формула (1) (или тем более формулы (2а), (2б)) применяются довольно редко. Все эти формулы тоже справедливы при любых (по величине и по знаку) значениях координат рассматриваемых точек и разностях этих координат, если только условиться, что угол α отсчитывается от A_1A_2 к A_1A_3 и может быть как положительным, так и отрицательным.

Так как равенство $\alpha=90°$ равносильно несуществованию величины $\operatorname{tg}\alpha$ (которая, так сказать, «обращается в бесконечность» — при этом знаменатель выражения для $\operatorname{tg}\alpha$ должен равняться нулю), то из формулы (1) вытекает следующее условие, необходимое и достаточное для того, чтобы прямые A_1A_2 и A_1A_3 были *взаимно перпендикулярны*:

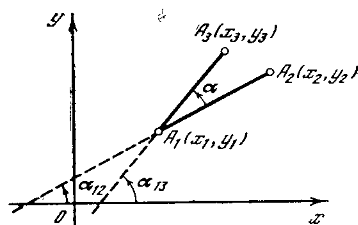


Рис. 1.9.1

$$(x_3-x_1)(x_2-x_1)+$$
$$+(y_3-y_1)(y_2-y_1)=0. \tag{1.9.3}$$

Из формул (1) и (2а), (2б) можно сделать и дальнейшие выводы. Заметим, что выражения, стоящие в знаменателях правых частей формул (2а) и (2б), равны расстояниям A_1A_2 и A_1A_3 (ср. с формулой (2.4)):

$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}=A_1A_2,$$
$$\sqrt{(x_3-x_1)^2+(y_3-y_1)^2}=A_1A_3.$$

А так как по известной формуле школьного курса геометрии

$$S_{\triangle A_1A_2A_3}=\frac{1}{2}A_1A_2\cdot A_1A_3\cdot|\sin\alpha|,$$

где $S_{\triangle A_1A_2A_3}$ — площадь треугольника $A_1A_2A_3$, то

$$S_{\triangle A_1A_2A_3}=$$
$$=\frac{1}{2}\,|(x_2-x_1)(y_3-y_1)-$$
$$-(x_3-x_1)(y_2-y_1)|. \tag{1.9.4}$$

Но, кроме того, $S_{\triangle A_1A_2A_3}=\frac{1}{2}A_1A_2\cdot h$, где h — расстояние от точки A_3 до пря-

мой A_1A_2. Поэтому в силу (4) имеем следующую формулу для расстояния[1] $h=h_{A_3,\,A_1A_2}$ от A_3 до A_1A_2:

$$h_{A_3,\,A_1A_2}=\frac{|(x_2-x_1)(y_3-y_1)-(x_3-x_1)(y_2-y_1)|}{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}. \qquad (1.9.5)$$

Отметим еще, что выбор системы координат на плоскости в геометрических задачах обычно является довольно произвольным: для задания координат надо указать на плоскости точку O (начало координат) и две проходящие через эту точку (взаимно перпендикулярные) оси координат Ox и Oy, которые, однако, можно выбирать по-разному.

Пусть, например, xOy и $x'O'y'$ — две разные системы координат (рис. 2). Для того чтобы установить связь между «старыми» координатами $(x,\ y)$ и «новыми» координатами $(x',\ y')$ одной и той же точки M, введем в рассмотрение еще и «промежуточную» систему координат x_1Oy_1, начало которой совпадает с началом старой системы координат, а оси параллельны осям новой системы. Чтобы связать координаты $(x,\ y)$ и $(x_1,\ y_1)$ какой-либо

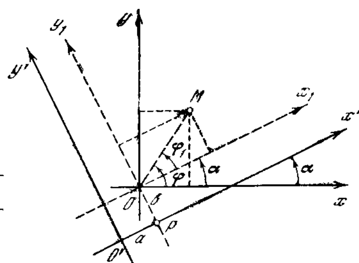


Рис. 1.9.2

точки, удобно рассмотреть также две системы полярных координат с одним и тем же полюсом O и полярными осями Ox и Ox_1 (см. рис. 2). Если $\widehat{xOx_1}=\alpha$, то точке $M(r,\ \varphi)$ (точке M, имеющей в первой системе полярных координат координаты $r,\ \varphi$) отвечают (полярные) координаты $r_1,\ \varphi_1$ во второй системе, где, очевидно, $r_1=r,\ \varphi_1=\varphi-\alpha$. Поэтому (см. формулы (2.3))

$$x=r\cos\varphi,\quad y=r\sin\varphi \quad\text{и}\quad x_1=r_1\cos\varphi_1=r\cos(\varphi-\alpha),\quad y_1=r\sin(\varphi-\alpha),$$

т. е.

$$x_1=r\cos(\varphi-\alpha)=r(\cos\varphi\cos\alpha+\sin\varphi\sin\alpha)=r\cos\varphi\cos\alpha+r\sin\varphi\sin\alpha=x\cos\alpha+y\sin\alpha,$$

$$y_1=r\sin(\varphi-\alpha)=r(\sin\varphi\cos\alpha-\cos\varphi\sin\alpha)=-r\cos\varphi\sin\alpha+r\sin\varphi\cos\alpha=-x\sin\alpha+y\cos\alpha.$$

С другой стороны, из того же рис. 2 следует, что

$$x'=x_1+a,\quad y'=y_1+b,$$

где $a=O'P$ и $b=PO$ — координаты старого начала O в новой системе координат $x'O'y'$.

Таким образом, окончательно получаем[2]

$$x'=x\cos\alpha+y\sin\alpha+a,$$
$$y'=-x\sin\alpha+y\cos\alpha+b. \qquad (1.9.6)$$

Формулы (6) играют в геометрии фундаментальную роль. Метод координат позволяет характеризовать каждую точку плоскости парой чисел $x,\ y$ — координатами этой точки; если у нас есть две точки, то им отвечают две пары чисел; множество точек можно заменить множеством пар чисел. Например, линии отвечает *однопараметрическое* множество пар $x(t)$, $y(t)$ чисел, зависящее от одного параметра t (ср. § 8). При этом переход к иной системе координат не отражается на геометрических свойствах фигур; поэтому смысл в геометрии имеют

---

[1] В формулах (4) и (5) можно опустить вертикальные черточки, считая площадь $S_{\triangle A_1A_2A_3}$ и расстояние $h_{A_3,\,A_1A_2}$ «направленными», т. е. имеющими тот или иной знак в зависимости от того, по какую сторону прямой A_1A_2 лежит точка A_3; здесь мы на этом останавливаться не будем.

[2] Мы здесь, как это принято в аналитической геометрии, ограничиваемся одними лишь правыми системами координат, т. е. такими, что поворот на 90° положительного луча Ox оси абсцисс, переводящий его в положительный луч Oy оси ординат, происходит в положительном направлении, т. е. в направлении, противоположном направлению вращения часовой стрелки. Нетрудно показать, что если xOy — правая система координат, а $x'Oy'$ — левая, то

$$x'=x\cos\alpha+y\sin\alpha+a,$$
$$y'=x\sin\alpha-y\cos\alpha+b$$

(какой смысл имеют здесь угол α и отрезки $a,\ b$?).

лишь такие соотношения между координатами точек, которые сохраняют свою форму при любом преобразовании (6) — при переходе от системы координат xOy к любой другой системе координат $x'O'y'$. Именно такими, разумеется, были все те соотношения, которые рассматривались выше.

По-другому сказанное можно пояснить так. Уравнение $y=f(x)$ или $F(x, y)$ некоторой линии Г зависит, очевидно, не только от самой линии, но и от выбора системы координат; оно описывает не линию Г (рис. 3, *а*), а гораздо более сложный образ: линию

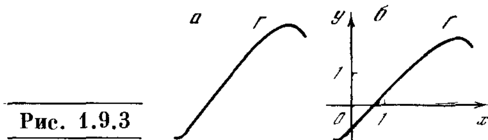


Рис. 1.9.3

Г и тот «координатный крест» из оси абсцисс и оси ординат, к которому отнесена линия (даже не только «координатный крест», но еще и единичные точки на осях, указывающие выбор единиц измерения величин x и y; см. рис. 3, *б*). Но геометрический смысл (а нас здесь интересует именно геометрия, а не физика[3]) имеют лишь такие связанные с уравнением кривой выражения, которые сохраняют свое значение при замене переменных по формулам (6). Так, например, в уравнении окружности с центром $Q(a, b)$ и радиусом r

$$(x-a)^2+(y-b)^2=r^2,$$

или

$$(x-a)^2+(y-b)^2-r^2=0, \qquad (1.9.7)$$

величины a и b характеризуют (всего лишь!) положение окружности на координатной плоскости, в то время как число r связано с «геометрией» кривой (с размерами окружности; см. ниже упр. 6).

Чтобы продемонстрировать сказанное, обратимся к формуле (2.4) для расстояния r_{12} между точками. Если $A_1(x_1, y_1)$ и $A_2(x_2, y_2)$ — две точки, то квадрат r_{12}^2 расстояния между ними, в силу (2.4) равен $(x_2-x_1)^2+(y_2-y_1)^2$.

[3] В физике обстоит дело не совсем так; см. по этому поводу § 9.8.

Введем теперь новые координаты x', y'; тогда согласно (6)

$$x_1'=x_1\cos\alpha+y_1\sin\alpha+a,$$
$$y_1'=-x_1\sin\alpha+y_1\cos\alpha+b,$$

и аналогично находятся новые координаты x_2', y_2' точки A_2 по ее старым координатам x_2, y_2. А теперь имеем

$$x_2'-x_1'=(x_2-x_1)\cos\alpha+(y_2-y_1)\sin\alpha,$$
$$y_2'-y_1'=-(x_2-x_1)\sin\alpha+(y_2-y_1)\cos\alpha$$

и соответственно

$$(x_2'-x_1')^2+(y_2'-y_1')^2=$$
$$=[(x_2-x_1)\cos\alpha+(y_2-y_1)\sin\alpha]^2+$$
$$+[-(x_2-x_1)\sin\alpha+(y_2-y_1)\cos\alpha]^2=$$
$$=(x_2-x_1)^2(\cos^2\alpha+\sin^2\alpha)+$$
$$+(y_2-y_1)^2(\sin^2\alpha+\cos^2\alpha)=$$
$$=(x_2-x_1)^2+(y_2-y_1)^2,$$

что и доказывает независимость величины r_{12} от выбора системы координат.

Ясно, что две прямые $y=k_1x+b_1$ и $y=k_2x+b_2$ (рис. 4) параллельны в том и только том случае, если их наклоны совпадают, т. е. если $k_1=k_2$.

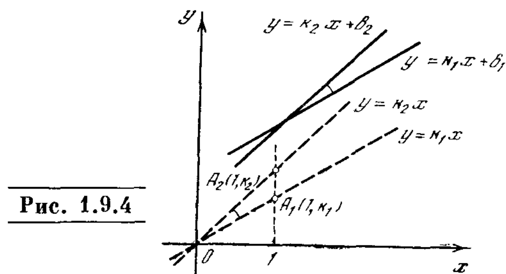


Рис. 1.9.4

Рассмотрим теперь проходящие через начало координат O прямые $y=k_1x$ и $y=k_2x$, параллельные нашим (вообще говоря, не параллельным) прямым (см. пунктирные линии на рис. 4). Поскольку этим новым прямым принадлежат точки $A_1(1,k_1)$ и $A_2(1,k_2)$, то в силу (1) угол α, равный углу A_1OA_2 между нашими прямыми (безразлично, старыми или новыми!), определяется по формуле

$$\operatorname{tg}\alpha=\frac{(k_2-0)(1-0)-(k_1-0)(1-0)}{(1-0)(1-0)+(k_2-0)(k_1-0)}$$

(ведь здесь роль точек $A_1(x_1, y_1)$, $A_2(x_2, y_2)$ и $A_3(x_3, y_3)$ формулы (1) играют точки $O(0, 0)$, $A_1(1, k_1)$ и

A_2 (1, k_2)). Таким образом, окончательно получаем

$$\operatorname{tg}\alpha=\frac{k_2-k_1}{1+k_2k_1}. \qquad (1.9.8)$$

В частности, две прямые $y=k_1x+b_1$ и $y=k_2x+b_2$ перпендикулярны тогда и только тогда, когда

$$k_1k_2+1=0 \quad \text{или} \quad k_1k_2=-1 \qquad (1.9.9)$$

(почему?).

Далее, пусть $y=kx+b$ и $y=kx+b_1$ — две параллельные прямые l и l_1, имеющие один и тот же наклон k и пересекающие ось y в точках $B(0,b)$ и $B_1(0,b_1)$ (рис. 5). В таком случае, очевидно,

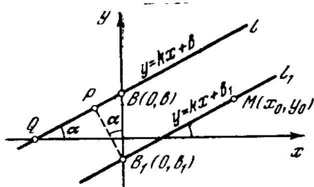


Рис. 1.9.5

$BB_1=|b_1-b|$; с другой стороны, перпендикуляр B_1P, опущенный из точки B_1 на прямую l, образует с осью y такой же угол α, какой образует прямая l (и l_1) с осью x (см. рис. 5, где $\angle BB_1P$ и $\angle xQl$ — углы со взаимно перпендикулярными сторонами). Но $\operatorname{tg}\widehat{xQl}=\operatorname{tg}\alpha=k$ по определению наклона прямой; поэтому и $\operatorname{tg}\widehat{BB_1P}=k$, и, значит,

$$\cos\widehat{BB_1P}=\cos\alpha=\frac{1}{\sqrt{1+\operatorname{tg}^2\alpha}}=\frac{1}{\sqrt{1+k^2}}.$$

А теперь из треугольника BB_1P (см. рис. 5) получаем

$$B_1P=BB_1\cos\alpha=\frac{|b_1-b|}{\sqrt{1+k^2}}.$$

Мы нашли, что расстояние $d=B_1P$ между параллельными прямыми l и l_1 равно

$$d=\frac{|b_1-b|}{\sqrt{1+k^2}}. \qquad (1.9.10)$$

Если l — прямая с уравнением $y=kx+b$, а $M(x_0, y_0)$ — произвольная точка (см. рис. 5), то параллельная l прямая l_1, проходящая через M, имеет уравнение

$$y-y_0=k(x-x_0), \text{ или } y=kx+b_1,$$

где $b_1=y_0-kx_0$ (см. выше формулу (3.3)). Поэтому в силу (10) расстояние h от точки M до прямой l равно

$$h=\frac{|y_0-kx_0-b|}{\sqrt{1+k^2}} \qquad (1.9.11)$$

(заметьте, что в числителе правой части (11) стоит абсолютная величина результата подстановки координат точки M в левую часть уравнения прямой $y-kx-b=0$).

Упражнения

1.9.1. Даны точка M и прямая l: а) $M=M(1, -1)$, $l: y=x+1$; б) $M=M(0, -1)$, l: $y=4$; в) $M=M(-2, 0)$, l: $y=-x-2$; г) $M=M(0, 0)$ (начало координат), l: $y=\frac{3}{4}x-\frac{1}{4}$. Проведите через точку M прямую p, перпендикулярную прямой l.

1.9.2. Даны три точки A_1, A_2, A_3: A_1 (0, 0), A_2 (4, 3), A_3 (−6, 8); A_1 (4, −4), A_2 (4, 0), A_3 (−1, 1); A_1 (1, 2), A_2 (2, 1), A_3 (−1, 1).

а) Найдите $\angle A_2A_1A_3$; б) вычислите $S_{\triangle A_1A_2A_3}$; в) чему равно $h_{A_1,\, A_2A_3}$?

1.9.3. Даны две параллельные прямые l и l_1: а) l: $y=x+2$, l_1: $y=x-2$; б) l: $3x-4y-1=0$, l_1: $-3x+4y=0$. Чему равно расстояние d между этими прямыми?

1.9.4. Для той же точки M и той же прямой l, что и в упр. 1, найдите расстояние h от M до l.

1.9.5. Пусть $A_1=A_1(x_1, y_1)$, $A_2=A_2(x_2,y_2)$, $A_3(x_3, y_3)$ и $A_4=A_4(x_4, y_4)$ — четыре точки; докажите, что $A_1A_2 \perp A_3A_4$ тогда и только тогда, когда

$$(x_2-x_1)(x_4-x_3)+(y_2-y_1)(y_4-y_3)=0.$$

1.9.6. Найдите, как преобразуется уравнение (7) окружности при переходе к новым координатам x', y' по формулам (6). Проверьте, что фигурирующие в (7) величины a и b меняют при этом свою величину (т. е. что новое уравнение будет иметь ту же форму, но с иными a и b), а величина r не меняется. [Указание. Для решения этой задачи удобно «обратить» формулы (6), выразив старые координаты x, y через новые координаты x', y'.]

1.9.7. Докажите, что при переходе (6) к другой системе координат сохраняют свою форму: а) условие $k_1=k_2$ параллельности двух прямых; б) условие (9) перпендикулярности двух прямых; в) формула (8) для угла между прямыми; г) формула (10) для расстояния между параллельными прямыми; д) формула (11) для расстояния от точки до прямой; е) формула (4) для площади треугольника; ж) формула (1) для угла треугольника; з) формула (5) для расстояния от точки до прямой; и) условие $A_1A_2 \perp A_3A_4$ (см. упр. 5). [См. указание к упр. 6.]

Глава 2

ЧТО ТАКОЕ ПРОИЗВОДНАЯ

§ 1. Движение, путь и скорость

Рассмотрим поступательное движение тела вдоль какой-то прямой линии; расстояние некоторой точки M тела от определенной точки O на этой прямой (координату точки M) обозначим через z, причем в одну сторону это расстояние будем считать положительным, а в другую — отрицательным. Пусть, например, прямая, вдоль которой движется тело, расположена вертикально; точки выше O соответствуют положительным z, точки ниже O — отрицательным z. Удобно представлять себе, что наше тело маленькое, так что можно просто говорить об одной «материальной точке» M и об ее расстоянии от определенной точки O прямой — от начала координат.

Задачи о движении тел с постоянной скоростью v приводят к простым арифметическим и алгебраическим расчетам (основанным на том, что путь равен произведению скорости на время, т. е. на элементарной формуле $s = vt$, где s —путь, а t — время). Однако в природе мы, как правило, имеем дело с движениями, скорость которых м е н я е т с я с течением времени. Исследование таких движений приводит к важным физическим понятиям *пути* и *скорости* как функций времени; математически уже здесь возникают основные понятия высшей математики — понятия *производной* и *интеграла*. В настоящей и следующей главах мы рассмотрим эту ситуацию более подробно.

Процесс движения материальной точки M заключается в том, что координата z этой точки меняется с изменением времени t. Движение тела (точки M) определяется зависимостью координаты z от времени t, т. е. оно характеризуется функцией $z=z(t)$. Зная функцию $z(t)$, можно найти положение z тела в любой момент времени t. Наглядное представление о движении дает график функции $z(t)$, для получения которого мы будем откладывать по оси абсцисс (ось t) время, а по оси ординат (ось z) — расстояние тела от начала отсчета O.

При *равномерном* движении тела (движение с постоянной скоростью v) путь s прямо пропорционален времени t, где коэффициент пропорциональности равен v (здесь $s=vt$). Обозначим координату тела в момент $t=0$ через z_0, тогда пройденный за время t путь s

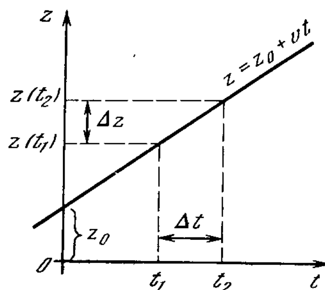


Рис. 2.1.1

будет равен разности $z(t)-z_0$. Таким образом, получаем

$$z(t) = z_0 + vt. \tag{2.1.1}$$

Мы видим, что при *равномерном* движении зависимость координаты z от времени t выражается *линейной* функцией. График зависимости $z(t)$ при равномерном движении представляет собой прямую линию на координатной (t, z)-плоскости (рис. 1).

При неравномерном движении зависимость $z(t)$ выражается более сложными формулами и соответствующий график представляет собой кривую линию. Рассмотрим теперь следующую основную задачу: *задана функция* $z(t)$, т. е. зависимость координаты тела от времени; *требуется найти скорость* v *движения тела*. В общем случае неравномерного движения скорость не остается постоянной: с течением времени она меняется. Значит, скорость v есть также функция времени t, и наша задача заключается в том, чтобы выразить эту неизвестную функцию $v(t)$ через известную функцию $z(t)$.

В частном случае равномерного движения все совсем просто. Скорость v определяется как путь, пройденный движущейся точкой за единицу времени. Найдем, например, путь, пройденный за 1 с: от момента t_1 до момента t_1+1 секунд. Этот путь (он численно равен скорости) измеряется разностью координат $z(t_1+1)$ и $z(t_1)$:

$$z(t_1+1)-z(t_1)=[z_0+v(t_1+1)]-$$
$$-[z_0+vt_1]=v.$$

Вместо этого можно также взять произвольный промежуток времени между моментами t_1 и t_2 и разделить пройденный путь z_2-z_1 на величину промежутка t_2-t_1:

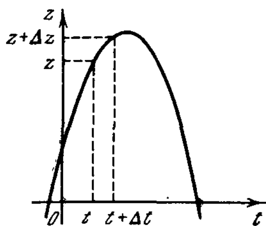


Рис. 2.1.2

$$\frac{z_2-z_1}{t_2-t_1}=\frac{(z_0+vt_2)-(z_0+vt_1)}{t_2-t_1}=v. \qquad (2.1.2)$$

Именно потому, что скорость постоянна, мы можем выбирать для ее вычисления любой интервал t_2-t_1; при этом получаемый ответ не будет зависеть ни от момента t_1, ни от величины интервала $|t_2-t_1|$. Однако в общем случае движения с переменной скоростью дело обстоит уже совсем не так.

Прежде чем переходить к общему случаю, полезно изменить обозначения. Назовем $t_1=t$, а $t_2=t+\Delta t$, так что разность t_2-t_1, т. е. рассматриваемый нами промежуток времени, мы обозначаем через Δt (см. рис. 1) [1]. Подобно этому путь

$$z(t_2)-z(t_1)=z(t+\Delta t)-z(t)=\Delta z.$$

В этих обозначениях формулу (2) можно переписать так:

$$\frac{\Delta z}{\Delta t}=v. \qquad (2.1.2а)$$

В общем случае движения с переменной скоростью правая часть (2а) выражает *среднюю скорость* движения на интервале времени от t до $t+\Delta t$ (интервал длительности Δt):

$$v_{ср}=\frac{\Delta z}{\Delta t}. \qquad (2.1.3)$$

Здесь приходится говорить о средней скорости потому, что сама скорость на протяжении интервала длительности Δt меняется.

Для примера положим, что путь z задается не линейной функцией, как на рис. 1, а квадратичной формулой

$$z(t)=z_0+bt+ct^2. \qquad (2.1.4)$$

На рис. 2 приведен один из возможных графиков, отвечающих функции вида (4). Вычислим по формуле (3) среднюю скорость $v_{ср}$ на интервале Δt. В нашем случае имеем

$$z(t)=z_0+bt+ct^2;$$
$$z(t+\Delta t)=z_0+b(t+\Delta t)+c(t+\Delta t)^2=$$
$$=z_0+bt+b\Delta t+ct^2+2ct\Delta t+c(\Delta t)^2$$

и, значит,

$$\Delta z=z(t+\Delta t)-z(t)=$$
$$=b\Delta t+2ct\Delta t+c(\Delta t)^2.$$

Отсюда

$$v_{ср}=\frac{\Delta z}{\Delta t}=b+2ct+c\Delta t. \qquad (2.1.5)$$

Сравним результаты (2а) и (5) для средней скорости при движении по закону (1) и по закону (4). Второй пример отличается тем, что в нем средняя скорость зависит и от самого момента t, и от промежутка времени Δt. Как же найти *мгновенную* скорость в момент t, зависящую единственно от этого момента?

Скорость меняется постепенно; поэтому чем меньше промежуток времени, на протяжении которого измеряется пройденный путь, тем меньше успеет измениться скорость, тем ближе будет средняя скорость к мгновенному ее значению. Формула (5) для $v_{ср}$ содержит два члена, не зависящих от величины промежутка Δt, и один член, пропор-

[1] Разумеется, Δ — это не множитель, а символ и Δt — не произведение Δ на t, подобно тому как $\sin\alpha$ не есть произведение sin на α: сокращать числитель и знаменатель правой части (2а) или (3) на Δ так же недопустимо, как нельзя сократить на sin числитель и знаменатель дроби $\frac{\sin\alpha}{\sin\beta}$. Символ Δ — это прописная греческая буква «дельта», заменяющая слово «приращение»; запись Δt (или Δz) читается как «дельта тэ» («дельта зет») или «приращение времени» («приращение пути»).

циональный Δt. При очень малых Δt и произведение $c\Delta t$ будет очень мало́. Поэтому при малых Δt последним членом в правой части (5) можно пренебречь: при этом формула для $v_{ср}$ обратится в формулу для мгновенной скорости:

$$v_{мгн} = b + 2ct. \qquad (2.1.6)$$

Так, например, пусть в формуле (4) $z_0=1$, $b=2$, $c=-1$; будем искать скорость движения при $t=0$. В таком случае (5) дает

$$v_{ср} = 2 + 2\cdot(-1)\cdot 0 + (-1)\cdot\Delta t = 2 - \Delta t.$$

Можно составить следующую таблицу:

Δt (от 0 до Δt с)	1	0,5	0,1	0,05	0,01
$v_{ср}$	1	1,5	1,9	1,95	1,99

Таким образом, чем меньше Δt, тем ближе выражение для $v_{ср}$ к 2; поэтому естественно считать, что $v_{мгн}=2$.

Внимательный читатель, вероятно, уже узнал в выражениях (4) и (6) известные из школьного учебника физики формулы для равноускоренного движения:

$$z(t) = z_0 + v_0 t + \frac{at^2}{2}, \qquad (2.1.7)$$

$$v(t) = v_0 + at;$$

только вместо коэффициента b формулы (4) в (7) стоит величина v_0 *начальной скорости* (скорости в момент $t=0$); коэффициент же c формулы (4) заменен на «полуускорение» $\frac{a}{2}$ (ибо a — *ускорение* движения).

Мы вычислили мгновенную скорость в момент времени t исходя из средней скорости в промежутке от t до $t+\Delta t$. Попробуем теперь вычислить ее, выбирая промежуток несколько по-иному. Найдем среднюю скорость в промежутке от $t_1 = t - \frac{3}{4}\Delta t$ до $t = t + \frac{1}{4}\Delta t$; длительность промежутка здесь по-прежнему равна Δt. В силу (4) имеем

$$z(t_1) = z_0 + b\left(t - \frac{3}{4}\Delta t\right) + c\left(t - \frac{3}{4}\Delta t\right)^2,$$

$$z(t_2) = z_0 + b\left(t + \frac{1}{4}\Delta t\right) + c\left(t + \frac{1}{4}\Delta t\right)^2$$

и

$$z(t_2) - z(t_1) = b\Delta t + 2ct\Delta t - \frac{1}{2}c(\Delta t)^2.$$

Отсюда следует, что

$$v_{ср} = \frac{z(t_2) - z(t_1)}{\Delta t} = b + 2ct - \frac{1}{2}c\Delta t. \qquad (2.1.8)$$

Если сравнить результаты (5) и (8), то видно, что средние скорости на интервалах от t до $t+\Delta t$ и от $t - \frac{3}{4}\Delta t$ до $t + \frac{1}{4}\Delta t$ отличаются на величину $c\Delta t\left[1 - \left(-\frac{1}{2}\right)\right] = \frac{3}{2}c\Delta t$. Но если мы хотим найти *мгновенную* скорость, то нужно брать очень маленький интервал Δt; тогда указанное различие между двумя выражениями для средней скорости уже не будет играть роли, и мы снова получим для мгновенной скорости прежнюю формулу $v_{мгн} = b + 2ct$. При этом, однако, величина (8) средней скорости на интервале от $t - \frac{3}{4}\Delta t$ до $t + \frac{1}{4}\Delta t$ ближе к мгновенной скорости (6), чем величина (5) средней скорости на интервале от t до $t+\Delta t$; поэтому при желании побыстрее (и поточнее) найти мгновенную скорость выбор интервала от t до $t+\Delta t$ менее удобен: здесь $v_{ср}$ дает худшее приближение для $v_{мгн}$. Еще лучше выбрать интервал так, чтобы интересующее нас значение t было серединой (центром) этого интервала; в случае квадратичной зависимости (4) между путем z и временем t полученное значение $v_{ср}$ будет точно совпадать с $v_{мгн}$ (см. упр. 2). Мы еще остановимся впоследствии на этом обстоятельстве (см. § 6.1).

Мы рассмотрели понятие мгновенной скорости для двух конкретных примеров: для равномерного и для равноускоренного движений. В § 3 мы дадим более точное определение мгновенной скорости при произвольном законе движения.

Упражнения

2.1.1. При движении материальной точки M по прямой наблюдалась следующая зависимость проходимого пути z от времени t: а) $z = t^3 + t$; б) $z = \frac{1}{1+t^2}$.

Чему равна средняя скорость $v_{ср}$ движения на интервале от момента времени t до $t+\Delta t$? Чему равна мгновенная скорость $v_{мгн}$ в момент t?

2.1.2. Для случая зависимости (4) пути от времени найдите величину $v_{ср}$ по формуле $v_{ср} = \frac{z(t + {}^1/_2\Delta t) - z(t - {}^1/_2\Delta t)}{\Delta t}$; сравните полученное выражение со значением $v_{мгн}$ в момент времени t.

§ 2*. Теплоемкость тела. Расширение тел при нагревании

Заметим, что процедура, аналогичная описанной в предыдущем параграфе, встречается во многих физических задачах. Здесь мы продемонстрируем это на двух простых примерах, апеллирующих лишь к тем сведениям, которые включены в школьный курс физики.

Напомним, что *теплоемкостью* того или иного вещества называется количество тепла (в джоулях), которое необходимо для нагревания 1 кг рассматриваемого вещества (воды, железа, золота и т. д.) на 1° С. Но ведь при различных начальных температурах для нагревания 1 кг вещества на 1° С требуется разное количество тепла, так что теплоемкость c вещества является функцией начальной температуры $T: c=c(T)$. Так, например, для нагревания 1 кг железа, взятого при температуре 0° С, на 1° С требуется 440, 857 Дж тепла, а для нагревания на 1° С того же количества железа, взятого при температуре 50° С, нужно уже 470, 583 Дж. Как же определить теплоемкость тела, отвечающую данной фиксированной температуре T?

Содержание § 1 подсказывает следующую методику определения интересующей нас величины $c=c(T)$ (теплоемкости при температуре T). Аналогом фигурирующего в § 1 пути z здесь, очевидно, будет количество Q тепла (в джоулях), которое надо придать 1 кг рассматриваемого вещества для нагревания его от исходной температуры (безразлично какой!) до температуры T; ясно, что эта величина зависит от T: $Q=Q(T)$. При этом для нагревания 1 кг вещества от температуры T_1 до температуры T_2 понадобится $Q(T_1, T_2)=Q(T_2)-Q(T_1)$ тепла; для нагревания тела от температуры T до $(T+\Delta T)$° С понадобится $Q(T+\Delta T)-Q(T)=\Delta Q$ тепла. Поэтому с р е д н я я теплоемкость $c_{ср}$ на участке от T до $(T+\Delta T)$, естественно, определится как

$$\frac{Q(T+\Delta T)-Q(T)}{\Delta T}=\frac{\Delta Q}{\Delta T}, \qquad (2.2.1)$$

а м г н о в е н н а я теплоемкость $c_{мгн}$ (прилагательное «мгновенная» в данном случае относится не к определенному моменту времени, а к фиксированной температуре T тепла) — как значение $c_{ср}$, отвечающее о ч е н ь м а л е н ь к о м у приращению ΔT температуры, причем полученное таким путем значение теплоемкости $c_{мгн}$ будет тем точнее, чем меньшее ΔT мы берем. Заметим, что в подавляющем большинстве случаев уже значение $\Delta T=1$° С будет достаточно ма̀ло для точного определения величины $c=c(T)$. Здесь выражение «достаточно мало́» означает, что полученное таким путем значение теплоемкости c практически не будет отличаться от значения, к которому мы придем, выбрав меньший (даже и гораздо меньший) интервал ΔT изменения температуры.

П р и м е р. Количество $Q=Q(T)$ тепла, необходимое для нагревания 1 кг железа от 0 до T° С, дается следующей эмпирически наблюдаемой зависимостью (вполне удовлетворительно передающей интересующий нас физический процесс, во всяком случае при $T<200$° С):

$$Q(T)=440{,}857\,T+0{,}29725\,T^2. \qquad (2.2.2)$$

Отсюда в полном соответствии с содержанием § 1 имеем (заметьте, что зависимость (2) Q от T, подобно зависимости (1.4), является квадратичной!):

$$c_{ср}=\frac{\Delta Q}{\Delta T}=$$

$$=\frac{440{,}857\,(T+\Delta T)+0{,}29725\,(T+\Delta T)^2}{\Delta T}-$$

$$-\frac{440{,}857T+0{,}29725T^2}{\Delta T}=$$

$$=440{,}857+0{,}5945T+0{,}29725\Delta T \qquad (2.2.3)$$

и, значит,

$$c_{мгн}=440{,}857+0{,}5945T. \qquad (2.2.4)$$

Таким образом, $c(0)=440{,}857$ Дж/(кг×град), как мы указывали выше, а, скажем, $c(100)=440{,}857+0{,}5945\times100=500{,}307$ Дж/(кг·град).

Аналогично обстоит дело и в других задачах, где надо определить величину, являющуюся характеристикой скорости изменения другой (меняющейся) величины. Так, например, *линейный коэффициент* k (теплового) *расширения* вещества обычно определяется как число, указывающее, насколько удлинится тонкий стержень из рассматриваемого вещества длиной в 1 см при нагревании его на 1° С. Однако и тут со-

ответствующее число k не является постоянным — оно зависит от начальной температуры T стержня; поэтому, строго говоря, наше определение, рассматривающее процесс нагревания стержня от T до $(T+1)°$ С, является не совсем корректным, поскольку в процессе этого нагревания интересующая нас величина $k=k(T)$ все время меняется.

Как же определить $k_{\text{мгн}}$, т. е. величину k, отвечающую строго фиксированному значению T температуры? Возьмем стержень длиной в 1 см при температуре 0° С и начнем его постепенно нагревать; при этом длина L стержня в процессе нагревания будет меняться, являясь функцией температуры: $L=L(T)$. При нагревании стержня от температуры T до температуры $(T+\Delta T)$, где приращение ΔT температуры уместно считать небольшим, длина стержня увеличится на величину

$$\Delta L = L(T+\Delta T) - L(T),$$

так что на рост температуры в 1° С приходится следующее среднее (или *удельное*) удлинение стержня:

$$\frac{L(T+\Delta T)-L(T)}{\Delta T} = \frac{\Delta L}{\Delta T}.$$

Однако это удлинение отнесено не к стержню длиной 1 см, а к стержню длиной L, так что на 1 см длины стержня приходится (опять же в среднем) удлинение

$$k_{\text{ср}} = \frac{1}{L}\frac{\Delta L}{\Delta T}. \qquad (2.2.5)$$

Задаваемая формулой (5) величина $k_{\text{ср}}$ и является приближением[1] для интересующего нас (мгновенного) коэффициента теплового удлинения вещества $k=k(T)$, причем это удлинение будет тем более точным, чем меньшее значение ΔT мы выбираем.

Пример. Для платины в широком интервале температур фигурирующая в наших рассмотрениях функция $L=L(T)$ хорошо приближается следующей эмпирической формулой:

$$L(T) = 1 + 8{,}806\cdot 10^{-6}T + 1{,}95\cdot 10^{-9}T^2 \qquad (2.2.6)$$

(снова квадратичная зависимость L от T). Выразим (линейный) коэффициент расширения $k=k(T)$ платины как функцию от T и найдем значения $k(0)$ и $k(1000)$.

Ответ. В точности, как выше (проделайте все расчеты самостоятельно!), находим, что

$$k(T) = \frac{1}{L}\frac{\Delta L}{\Delta T} \approx \approx \frac{8{,}806\cdot 10^{-6} + 3{,}90\cdot 10^{-9}T}{1+8{,}806\cdot 10^{-6}T + 1{,}95\cdot 10^{-9}T^2}, \qquad (2.2.7)$$

откуда следует, что

$k(0)=8{,}806\cdot 10^{-6}$ (1/°C) и $k(1000)= 12{,}57\cdot 10^{-6}$ (1/°C).

Упражнения

2.2.1. Для алмаза количество $Q(T)$ тепла (в Дж), необходимое для нагревания 1 кг вещества от 0 до T° С, в пределах от 0 до 700—800° С хорошо передается следующей эмпирической формулой:

$Q(T)=0{,}3965T+2{,}081\cdot 10^{-3}T^2-5{,}024\cdot 10^{-7}T^3$.

Найдите среднюю теплоемкость $c_{\text{ср}}$ алмаза на интервале от T до $(T+\Delta T)°$ С, а также (мгновенную) теплоемкость $c=c(T)$ как функцию T. Чему равна для алмаза величина $c(0)$, $c(100)$ и $c(500)$?

2.2.2. Количество $Q(T)$ тепла, необходимое для нагревания 1 кг воды от 0 до T° С, хорошо передается следующей эмпирической формулой:

$Q(T)=4186{,}68T+8373{,}36\cdot 10^{-5}T^2+ +1256\cdot 10^{-6}T^3$.

Чему равна теплоемкость $c(T)$ воды? Какое значение принимает она при $T=10°$, 90°?

2.2.3. Докажите, что (для любого вещества и при любой температуре T) *объемный* коэффициент теплового расширения (т. е. отнесенная к единице объема «мгновенная скорость» увеличения объема при росте температуры) равен утроенному линейному коэффициенту теплового расширения.

---

[1] Приближенный характер формулы (5) отражается и в том, что мы относим удлинение ΔL $\left(\text{точнее, } \textit{удельное удлинение } \frac{\Delta L}{\Delta T}\right)$ к длине $L=L(T)$ стержня в начальный момент времени, хотя реально длина в рассматриваемом процессе все время меняется (почему, собственно, мы делим отношение $\frac{\Delta L}{\Delta T}$ на $L(T)$, а не на $L(T+\Delta T)$?). Однако эта неточность в формуле (5) для k является весьма несущественной (почему?).

§ 3. Производная. Простейшие примеры вычисления производных

В § 1 в связи с задачей о мгновенной скорости движения мы пришли к рассмотрению отношений вида $\frac{z(t_2)-z(t_1)}{t_2-t_1}$, где t_2 и t_1 следует считать очень близкими между собой. Однако употребленное нами выражение «очень близкие» является «нематематическим», неопределенным, нестрогим. Точная формулировка тут такова: нас интересует *предел*, к которому стремится отношение

$$\frac{z(t_2)-z(t_1)}{t_2-t_1} \qquad (2.3.1)$$

при t_2, стремящемся к t_1. Используя обозначения Δt и Δz, это отношение можно переписать в виде

$$v_{\text{ср}}=\frac{\Delta z}{\Delta t}; \qquad (2.3.2)$$

условие $t_2 \to t_1$ принимает теперь вид $\Delta t \to 0$.

В формуле (2) величины Δt и Δz зависимы: можно выбрать любой промежуток времени $\Delta t=t_2-t_1$, но после того как промежуток времени Δt, стоящий в знаменателе, выбран, запись Δz в числителе правой части (2) означает уже не произвольный отрезок пути, а именно тот, который соответствует рассматриваемому промежутку времени Δt. В формуле (1) это следует из самой записи отношения, где в числителе стоит разность $z(t_2)-z(t_1)$ значений функции z при значениях t_2 и t_1 аргумента; правая часть (2) — это просто сокращенная запись отношения (1).

Итак, интересующая нас величина мгновенной скорости $v_{\text{мгн}}=v(t_1)$ в момент времени t_1 есть предел отношения $\frac{\Delta z}{\Delta t}$ при Δt, стремящемся к нулю (где $\Delta t=t_2-t_1$). Записывается это так:

$$v(t_1)=\lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t}.$$

Здесь $v(t_1)$ есть именно мгновенная скорость; буквы lim (начальные буквы латинского слова limes — «лимес», т. е. *предел*) обозначают предел; под ним записано, о каком именно пределе идет речь — при Δt, стремящемся к нулю; стрелка внизу заменяет слово «стремится», а справа указана та величина $\frac{\Delta z}{\Delta t}$, предел которой ищется. Читается запись $\lim\limits_{\Delta t\to 0}\frac{\Delta z}{\Delta t}$ обычно так: «предел дельта-зет по дельта-тэ при дельта-тэ, стремящемся к нулю», где слово «по» («дельта-зет по дельта-тэ») заменяет длинное выражение «деленное на соответствующее значение».

Что значит «предел», «стремление к пределу»? Те расчеты, которые мы произвели в § 1, служат наглядным пояснением этих понятий. Мы видели, что при малых промежутках Δt величина $v_{\text{ср}}$ во втором примере из § 1 отличалась от значения $v_{\text{мгн}}$ на малую величину, пропорциональную Δt. При малых Δt эта величина тоже будет мала: чем меньше Δt, тем она будет меньше, и при исчезающе малых Δt (т. е. таких, которые на фоне фигурирующих в (1.5) величин b и $2ct$ можно считать совсем незаметными) рассматриваемая величина тоже становится исчезающе малой. Именно поэтому при малых значениях Δt мы всегда можем пренебречь членом с Δt в выражении для $v_{\text{ср}}$.

Итак, отношение

$$\frac{\Delta z}{\Delta t}=\frac{z(t_2)-z(t_1)}{t_2-t_1}, \qquad (2.3.3)$$

или, как мы преимущественно писали в § 1, заменяя t_1 на t, а t_2 на $t+\Delta t$,

$$\frac{\Delta z}{\Delta t}=\frac{z(t+\Delta t)-z(t)}{\Delta t}, \qquad (2.3.3а)$$

стремится к определенному пределу при Δt, стремящемся к нулю[1]. Соответствующий предел есть мгновенная скорость v; он сам также является функцией t:

$$\lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t}=v(t). \qquad (2.3.4)$$

Почему при вычислении скорости по заданному выражению для $z(t)$ приходится проводить такой длинный расчет, находить Δz для различных Δt и только затем искать предел $\lim\limits_{\Delta t\to 0}\frac{\Delta z}{\Delta t}$?

---

[1] Можно также исходить из первоначальной формы (3) отношения $\frac{\Delta z}{\Delta t}$ и считать, что t_2 и t_1 стремятся к одному и тому же значению t, так что $\Delta t=t_2-t_1$ стремится к нулю (ср., в частности, текст, напечатанный мелким шрифтом в конце § 6.1).

Нельзя ли сразу подставить в наше отношение значение $\Delta t=0$? Нет, потому что при этом мы получили бы $\Delta z=0$: ведь $\Delta z=z(t+\Delta t)-z(t)$, и если $\Delta t=0$, то и $\Delta z=0$. Значит, при таком механическом способе действий мы пришли бы к дроби $\frac{\Delta z}{\Delta t}=\frac{0}{0}$, т. е. не получили бы никакого определенного значения этой дроби.

При вычислении скорости вся суть расчета заключается в том, чтобы брать малые Δt и соответствующие им малые Δz. При этом получается каждый раз вполне определенное отношение $\frac{\Delta z}{\Delta t}$; когда Δt уменьшается, стремится к нулю, то Δz уменьшается приблизительно пропорционально величине Δt, а поэтому отношение $\frac{\Delta z}{\Delta t}$ остается приблизительно постоянным: отношение $\frac{\Delta z}{\Delta t}$ стремится к определенному пределу при стремлении Δt к нулю [2]. Величина этого предела — это и есть мгновенная скорость $v(t)$ в случае, когда t — время, а z — путь.

Предел, к которому стремится отношение приращения функции к приращению независимой переменной при стремлении к нулю приращения независимой переменной, имеет первостепенное значение и для самой математики, и для многих ее приложений: выше мы видели, например, что такое важнейшее понятие, как (мгновенная) скорость движения, находится с помощью подобного предела. (В § 2 мы, по существу, сводили к аналогичному пределу задачу о вычислении (удельной) теплоемкости тела и о линейном коэффициенте теплового расширения вещества; продумайте самостоятельно эти примеры.) Поэтому предел рассматриваемого отношения имеет специальное название: *производная функция*, или, короче, просто ***производная***. Это название связано со следующим обстоятельством. Если z есть функция t, то предел (4) зависит как от вида функции $z(t)$, так и от того значения аргумента t, при котором вычисляется этот предел, т. е. этот предел также есть функция t — новая функция, которая задается (или порождается, или производится) функцией $z(t)$. Эту новую функцию естественно назвать производной функцией, где прилагательное «производная» подчеркивает ее зависимость от исходной, или основной, функции $z(t)$.

Для производной имеются специальные обозначения. Вот одно из них:

$$\frac{dz}{dt}\left(=\lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t}\right).$$

При этом величина $\frac{dz}{dt}$ (читается: «де-зет по де-тэ») рассматривается не как дробь, а просто как сокращенная запись предела, стоящего справа. Величина $\frac{dz}{dt}$ записана в форме дроби лишь для того, чтобы напомнить, что она получена из дроби $\frac{\Delta z}{\Delta t}$ путем перехода к пределу [3].

Другое обозначение производной — с помощью штриха: $v=z'(t)$ или, например для функции $y(x)$,

$$y'=y'(x)=\frac{dy}{dx}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}.$$

Иногда вместо знака функции подставляют ее выражение: так, например, если $z=at^2+b$, то можно вместо $\frac{dz}{dt}$ писать прямо $\frac{d(at^2+b)}{dt}$, или $(at^2+b)'$.

Итак, — и это очень важно — *производная функции определяется как предел отношения приращения функции к приращению независимой переменной при стремлении к нулю приращения независимой переменной*:

$$\frac{dy}{dx}=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}. \tag{2.3.5}$$

Мгновенная скорость движения тела равна производной координаты тела по времени. По аналогии также и в том случае, когда x не является временем, а y не имеет характера расстояния, говорят, что производная $\frac{dy}{dx}$ указывает

[2] При некоторых «неестественных» законах движения $z=z(t)$ этот предел может и не существовать; соответствующие оговорки будут сделаны ниже. Рекомендуем начинающему читателю пока не задумываться над этими исключениями.

[3] Ниже (см. § 4.1) мы увидим, что выражение $\frac{dz}{dt}$ $\left(\text{или } \frac{dy}{dx}\right)$ можно также понимать как дробь; однако пока читатель должен воспринимать эту запись слитно, как один символ.

скорость изменения функции y при изменении независимой переменной (аргумента) x. Так, например, в обозначениях рис. 1 можно сказать, что отношение $\frac{\Delta y}{\Delta x}$ является «средней скоростью» роста функции y на интервале AB изменения аргумента x, а предел этого отношения при B, стремящемся к A, выражает скорость роста y в точке A оси x.

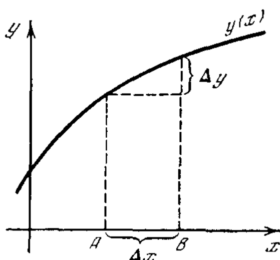


Рис. 2.3.1

Найдем теперь алгебраически производную от функции

$$z = t^2 \tag{2.3.6}$$

(по существу, мы это уже делали в § 1). Составим отношение

$$\frac{\Delta z}{\Delta t} = \frac{(t+\Delta t)^2 - t^2}{\Delta t}.$$

Раскроем скобки в числителе:

$$\Delta z = (t+\Delta t)^2 - t^2 = t^2 + 2t\Delta t + (\Delta t)^2 - t^2 = 2t\Delta t + (\Delta t)^2;$$

тогда получим

$$\frac{\Delta z}{\Delta t} = \frac{2t\Delta t + (\Delta t)^2}{\Delta t} = 2t + \Delta t. \tag{2.3.7}$$

Поскольку первое слагаемое в правой части (7) не зависит от Δt, то при стремлении Δt к нулю у нас, естественно, останется одно только это неубывающее в нашем процессе слагаемое:

$$\frac{dz}{dt} = \frac{d(t^2)}{dt} = \lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = \lim_{\Delta t \to 0} (2t + \Delta t) = 2t. \tag{2.3.6a}$$

Рассмотрим еще один пример:

$$z = t^3. \tag{2.3.8}$$

Здесь

$$\Delta z = (t+\Delta t)^3 - t^3 = t^3 + 3t^2\Delta t + 3t(\Delta t)^2 + (\Delta t)^3 - t^3 = 3t^2\Delta t + 3t(\Delta t)^2 + (\Delta t)^3,$$

$$\frac{\Delta z}{\Delta t} = 3t^2 + 3t\Delta t + (\Delta t)^2$$

и

$$\frac{dz}{dt} = \frac{d(t^3)}{dt} = \lim_{\Delta t \to 0} [3t^2 + 3t\Delta t + (\Delta t)^2] = 3t^2. \tag{2.3.8a}$$

В этих примерах предел можно было легко найти, так как при вычислении отношения $\frac{\Delta z}{\Delta t}$ величина Δt сокращалась. Но вот несколько более сложный пример:

$$z = \frac{1}{t}. \tag{2.3.9}$$

В этом случае

$$\frac{\Delta z}{\Delta t} = \frac{\frac{1}{t+\Delta t} - \frac{1}{t}}{\Delta t}.$$

Можно ли, когда мы перейдем к пределу, пренебречь в числителе величиной Δt в выражении $\frac{1}{t+\Delta t}$? Нет, потому что еще не проведено сокращение с величиной Δt в знаменателе. Заменяя $\frac{1}{1+\Delta t}$ на $\frac{1}{t}$, мы при малом Δt совершаем малую ошибку в одном из слагаемых числителя дроби $\frac{\Delta z}{\Delta t}$. Однако у этой дроби при малом Δt малы и числитель и знаменатель. Поэтому малая ошибка в числителе (в нашем случае она обращает числитель в нуль) недопустима.

Покажем правильный способ действий:

$$\Delta z = \frac{1}{t+\Delta t} - \frac{1}{t} = \frac{t-(t+\Delta t)}{t(t+\Delta t)} = -\frac{\Delta t}{t(t+\Delta t)},$$

$$\frac{\Delta z}{\Delta t} = -\frac{1}{t(t+\Delta t)}.$$

Теперь можно найти предел (производную), опуская Δt в знаменателе:

$$\frac{dz}{dt} = \frac{d(1/t)}{dt} = \lim_{\Delta t \to 0}\left[-\frac{1}{t(t+\Delta t)}\right] = -\frac{1}{t^2}. \tag{2.3.9a}$$

На этих примерах можно увидеть очень важное, основное свойство предела (собственно, это свойство принимают обычно за **определение** предела). При уменьшении величины Δt разность между значением отношения $\frac{\Delta z}{\Delta t}$ и пределом этого отношения (равным производной) $\lim_{\Delta t \to 0} \frac{\Delta z}{\Delta t} = \frac{dz}{dt}$ можно сде-

лать, как говорят, «сколь угодно малой», т. е. меньше любого заданного числа. При этом существенно, что далее в процессе изменения рассматриваемой величины $\left(\text{отношения } \frac{\Delta z}{\Delta t}\right)$ эта разность уже все время остается меньше выбранного нами числа.

Поясним это примером. Для $z=\frac{1}{t}$ имеем

$$\frac{dz}{dt}=-\frac{1}{t^2}, \quad \frac{\Delta z}{\Delta t}=-\frac{1}{t(t+\Delta t)}.$$

Возьмем, скажем, $t=2$, тогда $\frac{dz}{dt}=-0{,}25$. Можно ли выбрать такое Δt, чтобы $\frac{\Delta z}{\Delta t}$ отличалось от своего предела $-0{,}25$ меньше, чем на $0{,}0025$? Это значит, что Δt нужно выбрать так, чтобы

$$\frac{\Delta z}{\Delta t}\left(=-\frac{1}{2(2+\Delta t)}=-\frac{1}{4+2\Delta t}\right)$$

лежало в пределах между $-0{,}25+0{,}0025=-0{,}2475$ и $-0{,}25-0{,}0025=-0{,}2525$. Но легко проверить, что это наверное будет так, если Δt будет по абсолютной величине меньше $0{,}02$ (причем при уменьшении Δt, начиная со значения $0{,}02$, мы никогда не получим разности, превышающей $0{,}0025$).

Точно так же обстоит дело и для других функций: стремление к пределу при $\Delta t \to 0$ означает возможность такого выбора Δt, при котором достигается (и далее уже никогда не теряется) любая наперед нами выбранная степень приближения к пределу.

Особенно просто находится производная в частном случае $z=t$: очевидно, при этом $\Delta z=\Delta t$, $\frac{\Delta z}{\Delta t}=1$, т. е. отношение $\frac{\Delta z}{\Delta t}$ равно 1 для любых (больших и малых) Δt, а значит, и в пределе. Итак,

$$\text{если } z=t, \quad \text{то } \frac{dz}{dt}=\frac{dt}{dt}=1. \qquad \textbf{(2.3.10)}$$

Наконец, постоянную величину $z=c$ тоже можно рассматривать как частный случай функции — графиком этой функции будет параллельная оси x прямая (см. рис. 1.3.6). В этом случае, очевидно, $\Delta z=0$ при любых Δt; поэтому имеет место правило:

$$\text{если } z=c, \quad \text{то } \frac{dz}{dt}=\frac{dc}{dt}=0. \qquad \textbf{(2.3.11)}$$

§ 4. Первые свойства производной. Приближенное вычисление значений функции с помощью производной

Укажем теперь некоторые общие свойства производных.

Если функцию умножить на постоянный множитель, то на этот же множитель умножится и ее производная. Так, например,

$$\text{если } z=3t^2, \text{ то } \frac{dz}{dt}=\frac{d(3t^2)}{dt}=3\frac{d(t^2)}{dt}=3\cdot 2t=6t.$$

В общем виде:

$$\text{если } z(t)=ay(t), \quad \text{то } \frac{dz}{dt}=a\frac{dy}{dt}. \qquad \textbf{(2.4.1)}$$

Очевидно также, что *производная суммы двух функций равна сумме производных этих функций*:

$$\text{если } z(t)=x(t)+y(t),$$

$$\text{то } \frac{dz}{dt}=\frac{dx}{dt}+\frac{dy}{dt}. \qquad \textbf{(2.4.2)}$$

Последнее правило без труда переносится на сумму трех, четырех и вообще любого числа функций.

Используя теперь оба наши правила вместе, получим, что *производная суммы нескольких функций, взятых с какими-то* (какими угодно!) *постоянными коэффициентами, равна сумме производных этих функций, взятых с теми же коэффициентами*:

$$\text{если } z(t)=ax(t)+by(t)+cu(t),$$

$$\text{то } \frac{dz}{dt}=a\frac{dx}{dt}+b\frac{dy}{dt}+c\frac{du}{dt}. \qquad \textbf{(2.4.3)}$$

Каждое из названных правил легко доказать, исходя непосредственно из определения производной: ведь все они верны для приращения $\Delta z=z(t+\Delta t)-z(t)$ функции $z(t)$ (при любом Δt!); поэтому они выполняются и для отношения $\frac{\Delta z}{\Delta t}$, и для предела последнего отношения, т. е. для производной $\frac{dz}{dt}$. Так, например, если $z(t)=ay(t)$ (где a постоянно), то

$$\Delta z=z(t+\Delta t)-z(t)=ay(t+\Delta t)-ay(t)=a[y(t+\Delta t)-y(t)]=a\Delta y,$$

и, значит,

$$\frac{\Delta z}{\Delta t}=a\frac{\Delta y}{\Delta t} \quad \text{при всех } \Delta t, \text{ т. е. } \frac{dz}{dt}=a\frac{dy}{dt}.$$

Если $z(t)=x(t)+y(t)$, то

$$\Delta z=z(t+\Delta t)-z(t)=[x(t+\Delta t)+ +y(t+\Delta t)]-[x(t)+y(t)]=[x(t+\Delta t)- -x(t)]+[y(t+\Delta t)-y(t)]=\Delta x+\Delta y$$

и

$$\frac{\Delta z}{\Delta t}=\frac{\Delta x+\Delta y}{\Delta t}=\frac{\Delta x}{\Delta t}+\frac{\Delta y}{\Delta t},$$

откуда $\frac{dz}{dt}=\frac{dx}{dt}+\frac{dy}{dt}$,

и т. д.

Найдем теперь производную м н о г о ч л е н а. Мы уже знаем, что

$$\frac{dc}{dt}=0,\quad \frac{dt}{dt}=1,\quad \frac{d(t^2)}{dt}=2t,$$

$$\frac{d(t^3)}{dt}=3t^2. \qquad (2.4.4)$$

Поэтому

$$\frac{d(a+bt+ct^2+et^3)}{dt}= =\frac{da}{dt}+b\frac{dt}{dt}+c\frac{dt^2}{dt}+e\frac{dt^3}{dt}= =b+2ct+3et^2. \qquad (2.4.5)$$

Так, в § 1 фигурировала функция $z(t)=z_0+bt+ct^2$ (см. (1.4)). В силу (5) (где теперь надо положить $a=z_0$, $e=0$ имеем

$$v(t)=\frac{dz}{dt}=b+2ct,$$

т. е. тот результат, к которому мы без обращения к общим свойствам производных пришли в § 1.

Техника нахождения производных (или, как чаще говорят, *техника дифференцирования*) будет подробно изложена в гл. 4.

Забегая вперед, можно отметить, что дифференцирование функций, заданных теми или иными формулами, оказывается сравнительно простым и легким делом, более легким, например, чем решение алгебраических уравнений. Формулы для производных нередко оказываются даже проще (или, во всяком случае, не сложнее), чем формулы для самих исходных функций. Так, например, если исходная функция является многочленом, то ее производная тоже многочлен, причем этот новый многочлен оказывается более простым, чем исходный, в том смысле, что он имеет более низкую степень (ср. выше пример с многочленом третьей степени, производная которого оказывается квадратичной функцией переменного; так же обстоит дело и в случае многочленов л ю б ы х степеней). Если функция является алгебраической дробью, то и ее производная является дробью. Если функция содержит корни (дробные степени), то и производная содержит их. Производные от тригонометрических функций также являются тригонометрическими функциями. А в некоторых случаях, например для логарифмической или обратных тригонометрических функций, производная оказывается функцией более простого типа, чем исходная (для логарифма и арктангенса — алгебраической дробью).

Для нахождения производных не надо ждать вдохновения; здесь не нужна никакая изобретательность, выдумка, озарение, ибо задача всегда решается просто педантичным применением ряда простых правил типа перечисленных выше. Другие правила и примеры их применения, как уже говорилось, будут даны в гл. 4.

Все рассматривавшиеся до сих пор функции задавались формулами. Но не следует думать, что это обязательно для того, чтобы существовала производная. Например, мы можем считать зависимость пути от времени известной из опыта, заданной нам в виде очень подробных таблиц. С помощью этих таблиц можно искать величину мгновенной скорости (т. е. производной) почти так же, как мы делали это выше, составляя отношения $\frac{\Delta z}{\Delta t}=\frac{z(t_2)-z(t_1)}{t_2-t_1}$ для разных значений Δt и наблюдая, как будет меняться это отношение при уменьшении Δt. Разумеется, здесь уже нельзя говорить о *пределе* отношения $\frac{\Delta z}{\Delta t}$ при Δt, стремящемся к нулю, ибо для функции, заданной таблицей своих значений, величина Δt никак не может быть сделана сколь угодно малой (она не может стать меньше «шага» таблицы); поэтому здесь нельзя найти точное значение производной. Однако в большинстве случаев из таблицы можно получить достаточно хорошую оценку величины $\frac{dz}{dt}$, которая существует для всех «хороших», или «гладких», функций, независимо от их происхождения и от того, как задаются эти

функции. Укажем еще, что «арифметическое вычисление производной» — нахождение значений отношения $\frac{\Delta z}{\Delta t}$ (или — для функции $y=f(x)$ — отношения $\frac{\Delta y}{\Delta x}$) для ряда уменьшающихся значений Δt (или Δx) и наблюдение за поведением этих отношений очень упрощаются при пользовании любым малым компьютером.

Производная $\frac{dz}{dt}$ определяется как предел отношения приращений $\frac{\Delta z}{\Delta t}$ при $\Delta t \to 0$. При любом отличном от нуля Δt отношение приращений $\frac{\Delta z}{\Delta t}$, вообще говоря, не равно производной $\frac{dz}{dt}$, однако оно *приближенно* равно $\frac{dz}{dt}$; и это приближение будет тем лучше, чем меньше величина Δt. Таким образом, имеет место приближенное равенство

$$\frac{\Delta z}{\Delta t} \approx \frac{dz}{dt} = z'(t),$$

$$\Delta z \approx \frac{dz}{dt}\Delta t = z'(t)\Delta t. \qquad (2.4.6)$$

Отсюда найдем и приближенное значение функции $z(t+\Delta t)$:

$$z(t+\Delta t) = z(t) + \Delta z \approx z(t) + \frac{dz}{dt}\Delta t = z(t) + z'(t)\Delta t. \qquad (2.4.7)$$

Обратите внимание на то, что в формуле (7) первое равенство является точным (ведь это просто определение Δz!), а второе — приближенным.

Вернемся к обозначениям $t_2=t+\Delta t$, $t_1=t$, которыми мы пользовались раньше. Тогда (7) перепишется так:

$$z(t_2) \approx z(t_1) + z'(t_1)(t_2-t_1). \qquad (2.4.7а)$$

Таким образом, при малой разности t_2-t_1, т. е. при t_2, близком к t_1, функция $z(t_2)$ может быть выражена приближенной формулой, в которую входят значения функции $z(t)$ и ее производной $z'(t)$ при $t=t_1$. Отметим, что в эту формулу t_2 входит в первой степени, или, как говорят еще, *линейно*.

Формула (7) (мы еще вернемся к ней в § 4.1) является очень важной, поэтому обсудим ее подробнее. Рассмотрим, например, функцию $v=x^3$ — для наглядности можно себе представить, что речь здесь идет об объеме v куба с ребром x. Производная $\frac{dv}{dx}$ в этом случае, как мы знаем, равна $3x^2$; таким образом, формула (7) конкретизируется здесь так:

$$v(x+\Delta x) = (x+\Delta x)^3 \approx x^3 + 3x^2\Delta x,$$

т. е.

$$\Delta v = v(x+\Delta x) - v(x) \approx 3x^2\Delta x, \qquad (2.4.8)$$

в то время как точное выражение для $v(x+\Delta x)$ в силу известной формулы имеет вид

$$v(x+\Delta x) = (x+\Delta x)^3 = x^3 + 3x^2\Delta x + 3x(\Delta x)^2 + (\Delta x)^3,$$

откуда

$$\Delta v = 3x^2\Delta x + 3x(\Delta x)^2 + (\Delta x)^3. \qquad (2.4.8а)$$

Пусть теперь сторона куба равна 1 м; тогда его объем, разумеется, равен 1 м^3. Что же можно сказать, если выяснится, что при измерении стороны куба мы допустили ошибку, что эта сторона несколько больше величины 1 м, скажем, на 1 или на 5 мм или даже на 1 или 2 см?

Если наша ошибка равна 1 мм, т. е. реально $x=1{,}001$ м, то первоначальное значение $v=1$ объема надо будет увеличить на величину Δv, даваемую формулой (8а). Оценим теперь вклад в стоящую в правой части (8а) сумму отдельных ее слагаемых:

$$3x^2\Delta x = 3\cdot 1^2\cdot 0{,}001 = 0{,}003,$$

$$3x(\Delta x)^2 = 3\cdot 1\cdot (0{,}001)^2 = 0{,}000003,$$

$$(\Delta x)^3 = (0{,}001)^3 = 0{,}000000001;$$

при этом

$$v(1{,}001) = 1{,}003003001,$$

$$\Delta v = 0{,}003003001. \qquad (2.4.9)$$

Но имеет ли смысл выписывать в записи числа v все эти цифры? Ведь 3-й член стоящей в правой части (8а) суммы меньше 1-го ее члена в 3 000 000 (три миллиона!) раз; поэтому сохранение здесь 3-го члена ничем нельзя оправдать. Но и 2-й член той же суммы меньше первого в 1000 раз, так что и он обеспечивает лишь чисто иллюзорную точность: ведь если ошибка в измерении ребра куба окажется равной не точно 1 мм, а, скажем, 1,1 мм (разумеется, — абсолютно реалистическое предположение!), то 1-й член суммы (8)

обратится в 0,0033 вместо первоначального его значения 0,003, и все остальные цифры записей (9) окажутся смазанными, совершенно недостоверными. Так зачем же мы будем производить бессмысленную работу и загромождать расчеты ничего не означающими цифрами?

Аналогичная картина наблюдается и при иных (но все еще достаточно малых) значениях Δx. Так, при $\Delta x = 0{,}005$ м ($= {}^1/_2$ см) имеем

$$3x^2\Delta x = 3\cdot 1^2\cdot 0{,}005 = 0{,}015,$$

$$3x(\Delta x)^2 = 3\cdot 1\cdot (0{,}005)^2 = 0{,}000075,$$

$$(\Delta x)^3 = (0{,}005)^3 = 0{,}000000125.$$

Таким образом, окончательно получаем

$$v\,(1{,}005) = (1{,}005)^3 = 1{,}015075125,$$

где тоже, конечно, вполне достаточно (и даже более того — как правило, необходимо) ограничиться приближением

$$v \approx 1{,}015.$$

Подробнее все связанные с этим примером численные расчеты просуммированы в следующей таблице (из которой видно, что еще при $\Delta x = 0{,}02 = 2$ см, а при невысоких требованиях к точности результата и при $\Delta x = 0{,}05 = 5$ см замена выражения $(x+\Delta x)^3$ на $x^3 + 3x^2\Delta x$ является вполне допустимой):

$x = 1+\Delta x$	$1+$ $+0=1$	1,001	1,005	1,01
$v = (1+\Delta x)^3$	1	1,003003	1,015075	1,0303
$1+3\Delta x$	1	1,003	1,015	1,03

$x = 1+\Delta x$	1,02	1,05	1,1	1,5	2
$v = (1+\Delta x)^3$	1,0612	1,1576	1,3310	3,375	8
$1+3\Delta x$	1,06	1,15	1,30	2,50	4

Вот еще один пример: пусть $y = \sqrt{x}$ (скажем, y — сторона квадратной пластины площади или массы[1] x). Найдем значения функции y при x, близком к 4. В этом случае $y\,(4) = \sqrt{4} = 2$. Производная $y'\,(x) = \frac{1}{2\sqrt{x}}$ (см. ниже упр. 2); поэтому $y'\,(4) = \frac{1}{2\sqrt{4}} = \frac{1}{4}$, — и приближенная формула (7) имеет вид

$$y\,(x) = \sqrt{4+\Delta x} \approx 2 + 0{,}25\Delta x.$$

Сравним и здесь приближенное и точное значения $y(x)$:

$x = 4+\Delta x$	4	4,1	4,5	5
$y = \sqrt{4+\Delta x}$	2	2,02485	2,1213	2,24
$2+0{,}25\Delta x$	2	2,025	2,125	2,25

$x = 4+\Delta x$	6	7	8	9
$y = \sqrt{4+\Delta x}$	2,45	2,65	2,83	3
$2+0{,}25\Delta x$	2,50	2,75	3,0	3,25

Вернемся опять к нашему основному примеру время—путь—скорость, т. е. представим себе, что t есть время, $z\,(t)$ — проходимый за время t путь, $z'\,(t) = \frac{dz}{dt}$ — мгновенная скорость, Δz — приращение пути, т. е. путь, пройденный за малое время Δt. Формула

$$\Delta z \approx z'\,(t)\,\Delta t \qquad (2.4.10)$$

означает, что путь приближенно равен произведению мгновенной скорости на промежуток времени. Но ведь мгновенная скорость сама меняется с течением времени; поэтому пользоваться приближенным равенством (10) можно лишь в том случае, когда мгновенная скорость не успела заметно измениться за время Δt. Значит, чем быстрее меняется величина $z'(t)$, тем меньшее Δt нужно брать в формуле (10). И наоборот, чем медленнее изменяется $z'(t)$, тем бо́льшим можно брать Δt, т. е. величина приращения Δt, для которого формула (10) все еще дает малую ошибку, зависит от скорости изменения производной на промежутке Δt. Разумеется, все это верно и для произвольной функции

[1] Масса m однородной квадратной пластины со стороной y равна ρy^2, где ρ — постоянное число (плотность пластины), но функция $m = \rho y^2$ несущественно отличается от функции $x = y^2$ (ср. со сказанным по этому поводу в § 1.7).

$y=f(x)$ какого угодно аргумента x, где также производная $\frac{dy}{dx}$ может рассматриваться как скорость изменения функции y при данном значении аргумента x (см. рис. 3.1 и относящийся к нему текст).

Рассмотренные выше примеры (где мы опять обозначаем аргумент через t, а значение функции — через z и принимаем t за время, а z за путь) подтверждают этот вывод. В первом примере, где $z=t^3$, при изменении t от 1 до 2 (при $\Delta t=1$) производная $z'(t)=3t^2$ изменяется от 3 до 12 (т. е. в 4 раза). Во втором примере $\left(z=\sqrt{t}\right)$ при изменении t от 4 до 9 (при $\Delta t=5$) производная $z'(t)=\frac{1}{2\sqrt{t}}$ изменяется от 0,25 до 0,167 (т. е. примерно на 30%). Поэтому приближенная формула (7) (или (10)) во втором случае дает хороший результат и при сравнительно больших значениях Δt: так, при $\Delta t=4$, что составляет 100% от первоначального значения t, ошибка при ее применении будет равна всего лишь 6% от истинного значения функции, а при $\Delta t=5$, где Δt равно 125% от исходного значения t, ошибка составляет $1/12\approx 8\%$ от истинного значения. Напротив, в случае зависимости $z=t^3$ использование приближения (7) совершенно недопустимо уже при $\Delta t=0{,}5$, когда ошибка составляет более 25% верного результата (при $\Delta t=1$ ошибка равна 100% результата). Разумеется, все сказанное в равной степени относится к случаям положительных и отрицательных Δt (ср. упр. 5 ниже).

Подробно вопрос об оценке ошибки в приближенной формуле (7) и об уточнениях этой формулы будет разобран в гл. 6.

Предваряя содержание первых параграфов гл. 6, укажем, что при малых разностях t_2-t_1 функцию $z(t_2)$ можно представить в виде

$$z(t_2)=z(t_1)+a(t_2-t_1)+b(t_2-t_1)^2+ \\ +c(t_2-t_1)^3+\ldots, \qquad (2.4.11)$$

причем равенство (11) можно даже считать точным, если только допустить, что справа стоит сумма б е с к о н е ч н о г о числа слагаемых (точный смысл этого будет разъяснен в гл. 6). Так как мы полагаем, что t_2-t_1 мало, то в правой части (11) каждый следующий член меньше предыдущего, ибо он содержит более высокую степень малой величины t_2-t_1. При этом первые два члена общей формулы (11) в точности совпадают с приближенной формулой (7а), ибо $a=z'(t_1)=\frac{dz(t_1)}{dt}$. Отличие приближенной формулы от точной связано лишь с наличием в точной формуле члена, пропорционального $(t_2-t_1)^2$, и следующих членов, содержащих еще более высокие степени этой малой величины.

Общепринята такая терминология: говорят, что (t_2-t_1) и $a(t_2-t_1)$ при малой разности $(t_2-t_1)=\Delta t$ — суть величины *первого порядка малости* (здесь и далее имеется в виду, что t_2 стремится к t_1 и Δt уменьшается неограниченно). Величины $(t_2-t_1)^2$, а также $b(t_2-t_1)^2$ называют величинами *второго порядка малости*, $(t_2-t_1)^3$ и $c(t_2-t_1)^3$ — величинами *третьего порядка малости* и т. д. Значит, формула (7а) верна с точностью до малых первого порядка, ошибка формулы (7а) — второго порядка малости.

Если вывести из (7а) приближенное выражение для производной, то нам придется разделить обе части (7а) на t_2-t_1; при этом порядок малости уменьшится. Равенство

$$z'(t_1)\approx\frac{z(t_2)-z(t_1)}{t_2-t_1}$$

является приближенным, причем ошибка в нем будет уже не второго, а первого порядка малости, т. е. пропорциональна t_2-t_1; но все равно при $t_2\to t_1$, т. е. при $\Delta t=t_2-t_1\to 0$ эта ошибка стремится к нулю. Именно это мы утверждали и раньше, определяя производную.

Упражнения

2.4.1. Найдите производные функций: а) $y=x^4$; б) $y=4x^3-3x^2+2x-1$; в) $y=(2x+1)^2$; г) $y=\frac{1}{x^2}$; д) $y=a\left(x+\frac{1}{x}\right)$; е) $y=ax^2+\frac{b}{x^2}$.

2.4.2. Докажите, что производная функции $y=\sqrt{x}$ равна $\frac{1}{2\sqrt{x}}$. [У к а з а н и е. Умножьте числитель и знаменатель выражения $\frac{\sqrt{x+\Delta x}-\sqrt{x}}{\Delta x}$ на сумму $\sqrt{x+\Delta x}+\sqrt{x}$.]

2.4.3. Докажите, что производная функции $y=\sqrt{x^3}$ равна $\frac{3}{2}\sqrt{x}$. [У к а з а н и е. Вос-

пользуйтесь приемом, аналогичным использованному в решении упр. 2.]

2.4.4. Найдите $(1{,}2)^2$, $(1{,}1)^2$, $(1{,}05)^2$, $(1{,}01)^2$, пользуясь формулой (7). Сравните полученные результаты с точными.

2.4.5. Для функции $z(t)=2+20t-5t^2$ найдите при помощи производной значения $z(1{,}1)$, $z(1{,}05)$, $z(0{,}98)$. Сравните полученные результаты с точными значениями. [Указание. В последнем случае примите $t=1$, $\Delta t=-0{,}02$.]

§ 5. Касательная к кривой

С помощью производной можно решить важную геометрическую задачу — задачу о нахождении *касательной* к кривой, заданной уравнением $y=f(x)$. Координаты точки касания A здесь считаются известными: $x=x_0$, $y=y_0=f(x_0)$.

С точки зрения аналитической (координатной) геометрии Декарта (см. гл. 1) найти касательную — это значит найти ее уравнение. Очевидно, что касательная — это одна из прямых, проходящих через точку касания. Но уравнение любой прямой, проходящей через заданную точку $A(x_0, y_0)$, имеет вид $y-y_0=k(x-x_0)$ (см. формулу (1.3.3)). Для того чтобы найти уравнение касательной, остается определить величину k — угловой коэффициент, или наклон, касательной (ее «крутизну»). Для этого сперва найдем наклон прямой, проходящей через две заданные точки A, B рассматриваемой кривой (рис. 1); такую прямую называют *секущей*.

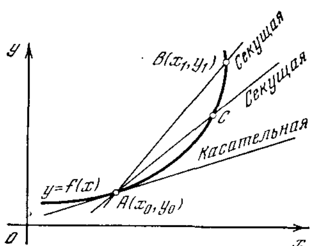


Рис. 2.5.1

Когда эти две точки кривой сближаются, секущая приближается к касательной.

На рис. 1 показаны две секущие, первая из которых проходит через точки A и B, а вторая — через точки A и C, где точка C расположена ближе к точке A, чем точка B. Чем ближе точка B или C к A, тем ближе секущая AB (или AC) к касательной. Поэтому наклон касательной равен п р е д е л у, к которому стремится наклон секущей при стремлении друг к другу точек кривой, через которые проходит секущая. (Это обстоятельство даже можно считать о п р е д е л е н и е м касательной; разумеется, точнее здесь было бы говорить, что точка B стремится к точке A или что две точки B, C кривой стремятся к A, а секущая BC — к касательной в этой выбранной точке A.)

Наклон $k_с$ секущей легко выразить через координаты точек пересечения секущей с кривой.

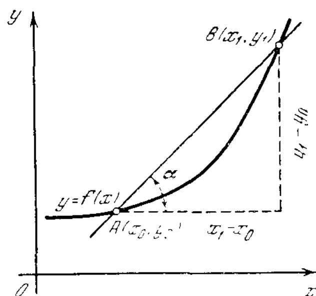


Рис. 2.5.2

Возьмем в качестве одной из этих точек ту точку $A(x_0, y_0)$, касательная в которой нас интересует; координаты второй точки B пересечения секущей с кривой обозначим через x_1, y_1. Так как обе эти точки принадлежат кривой с уравнением $y=f(x)$ (принадлежат линии $y=f(x)$, как обычно говорят), то $y_0=f(x_0)$ и $y_1=f(x_1)$. Как видно из рис. 2, наклон $k_с$ секущей равен

$$k_с=\operatorname{tg}\alpha=\frac{y_1-y_0}{x_1-x_0}=\frac{f(x_1)-f(x_0)}{x_1-x_0}$$

(см. § 1.3).

Чтобы вычислить наклон касательной в точке $x=x_0$, нужно брать точку B все ближе к A, т. е. нужно, чтобы x_1 стремилось к x_0. Следовательно, наклон k касательной равен пределу величины $k_с$ при x_1, стремящемся к x_0:

$$k=\lim_{x_1\to x_0} k_с=\lim_{x_1\to x_0}\frac{f(x_1)-f(x_0)}{x_1-x_0}.$$

Обозначим разность x_1-x_0 через Δx; тогда $x_1=x_0+\Delta x$ и $\Delta f=f(x_1)-f(x_0)=f(x_0+\Delta x)-f(x_0)$. В этих обозначениях наклоны $k_с$ секущей и k касательной выразятся формулами:

$$k_с=\frac{\Delta f}{\Delta x},\quad k=\lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x}.$$

Следовательно, *угловой коэффициент (наклон) касательной равен производной функции $f(x)$ в точке x_0*:

$$k=\frac{df}{dx}=f'(x_0). \qquad (2.5.1)$$

Мы знаем, что производная $f'(x)$ функции $f(x)$ сама является функцией x. Так как мы искали наклон касательной в фиксированной точке $A(x_0, y_0)$, то при вычислении предела отношения $\frac{\Delta f}{\Delta x}$ мы считали закрепленным значение $x=x_0$. Поэтому в окончательной формуле и стоит $f'(x_0)$ — значение производной функции $f'(x)$ при $x=x_0$. Переменная же функция $f'(x)$, очевидно, выражает наклон касательной к кривой $y=f(x)$ в переменной точке (x, y) (или $(x, f(x))$) этой кривой.

Рассмотрим пример. Пусть наша кривая является параболой $y=x^2$, т. е. пусть $f(x)=x^2$. Составим уравнение касательной к параболе в точке $x_0=2$, $y_0=f(x_0)=4$.

Мы знаем производную функции x^2:

$$f'(x)=\frac{df}{dx}=\frac{dx^2}{dx}=2x$$

(ср. (3.6а)). Следовательно, в интересующей нас точке наклон касательной

$$k=f'(x_0)=2x_0=4,$$

а уравнение касательной (рис. 3) имеет вид

$y-y_0=k(x-x_0)$, т. е. $y-4=4(x-2)$, $y=4x-4$.

Вообще наклон касательной к параболе $y=x^2$ в точке (x_0, x_0^2) равен $2x_0$ $\left(\text{ибо здесь } \frac{dy}{dx}=2x\right)$; следовательно, уравнение касательной имеет вид

$$y-x_0^2=2x_0(x-x_0), \text{ или } y=2x_0x-x_0^2. \quad (2.5.2)$$

Тем самым мы вновь получили результат, установленный в § 1.4 совсем другим путем: прямая $y=kx+b$ в том и только в том случае касается параболы $y=x^2$, если коэффициенты k и b уравнения прямой можно представить как выражения $2x_0$ и $-x_0^2$, зависящие от неопределенного числа (параметра) x_0; другими словами, условие касания прямой $y=kx+b$ с параболой $y=x^2$ имеет вид $b=-(k/2)^2$ (ср. (1.4.8а)).

Вот еще пример: рассмотрим полукубическую параболу $y=\sqrt{x^3}$ (см. § 1.5). Здесь $y'=\frac{3}{2}\sqrt{x}$ (см. упр. 4.3); поэтому $k=f'(x_0)=\frac{3}{2}\sqrt{x_0}$. В частности, в начале координат O наклон касательной равен $k=f'(0)=\frac{3}{2}\sqrt{0}=0$; поэтому в точке O полукубическая парабола касается оси x, как и изображено на рис. 1.5.7.

Без помощи производных трудно провести касательную к кривой, заданной уравнением $y=f(x)$: нужно вычислить координаты большого числа точек кривой, с помощью лекала провести кривую по этим точкам и потом

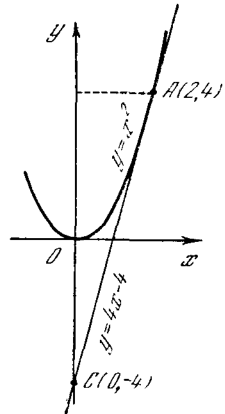


Рис. 2.5.3

на глаз приставлять линейку к кривой в заданной точке, внимательно следя за тем, чтобы линейка касалась, а не пересекала кривую. С помощью производных мы находим уравнение касательной, по этому уравнению находим две точки, принадлежащие заданной этим уравнением прямой, и затем с помощью линейки проводим прямую — касательную, проходящую через эти две точки. В качестве одной точки естественно взять саму точку касания $A(x_0, y_0)$, вторую точку C на прямой лучше взять далеко от A, тогда с большой точностью определится наклон и положение касательной — прямой, проходящей через известные точки A и C.

Так, например, выше мы установили, что уравнение прямой, касающейся параболы $y=x^2$ в точке $x_0=2$, $y_0=4$, имеет вид $y=4x-4$. Найдем координаты двух точек этой прямой; при $x=2$ мы имеем $y=4$; это есть сама точка касания $A(2, 4)$ (ее координаты можно было и не вычислять — ведь касательная обязана пройти через нее). В качестве второй точки (C) выберем точку пересечения касательной с осью y: положим $x=0$, найдем $y=-4$, так что $C=C(0, -4)$ (см. рис. 3).

Отметим любопытное обстоятельство: *при $x=0$ $y=-y_0$*, т. е. *точка C пересечения касательной с осью y лежит на столько ниже оси x, на сколько сама точка касания лежит выше оси x.* Это не случайно: наше правило справедливо для всех касательных к параболам с урав-

нением вида $y=ax^2$ (при любом $a>0$). В самом деле, если касательная проведена в точке $A(x_0,\ y_0=ax_0^2)$, то ее уравнение имеет вид

$$y-y_0=2ax_0(x-x_0) \tag{2.5.2a}$$

(ср. с равенством (2)), и при $x=0$ мы получаем $y-y_0=-2ax_0^2$, $y=y_0-2ax_0^2=y_0-2y_0=-y_0$. Таким образом, касательная к параболе проходит через точки $A(x_0,\ y_0=ax_0^2)$ и $C(0,\ y=-y_0=-ax_0^2)$.

При построении кривой по точкам трудно точно провести кривую, если вычислено мало точек. С помощью производных можно заранее провести ка-

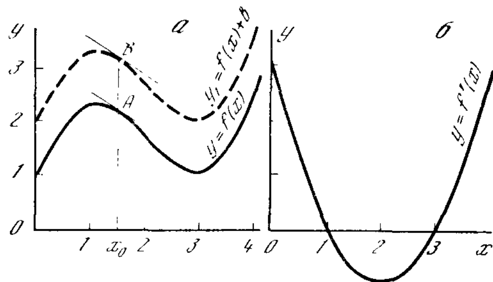


Рис. 2.5.4

сательные к кривой в этих точках; после этого можно точнее провести саму кривую.

Наглядно ясно, что в точках, в которых функция принимает наибольшее или наименьшее значение — в точках максимума и минимума функции, — касательная к графику функции горизонтальна; ниже мы еще не раз вернемся к этому важному обстоятельству. Уравнение горизонтальной прямой имеет вид $y=\text{const}$; наклон горизонтальной прямой $k=0$.

Таким образом, мы приходим к так называемой теореме Ферма: *в точках максимума и минимума кривой, являющейся графиком функции $y=f(x)$, производная функции $f(x)$ равна нулю.*

С помощью этой теоремы можно находить координаты x точек минимума и максимума кривой; соответствующие координаты y при этом легко найти, подставляя найденные значения x в уравнение кривой. Ясно, что, зная координаты точек максимума и минимума, можно точнее провести саму кривую.

Полезно в порядке упражнения, нарисовав от руки кривую $y(x)$, хотя бы приблизительно, но быстро провести кривую $y'(x)$, обращая внимание на знак $y'(x)$ и на точки, где $y'(x)$ обращается в нуль. Такой пример показан на рис. 4, *а* (график $y(x)$) и рис. 4, *б* (график производной $y'(x)$).

Точки обращения в нуль самой функции $y(x)$ для производной $y'(x)$ ничем не примечательны. Если кривую $y(x)$ переместить параллельно самой себе (верхняя кривая на рис. 4, *а*), т. е. сдвинуть вверх (или вниз) на любой отрезок b, то кривая $y'(x)$ от этого никак не изменится: при параллельном переносе (сдвиге) кривой в вертикальном направлении наклон любого участка кривой (точнее, следовало бы говорить о наклоне касательной к кривой в любой ее точке) останется тем же самым (ср. верхнюю и нижнюю кривые на рис. 4, *а* и, в частности, касательные к этим кривым в точках A и B). Это полностью соответствует тому, что функции $y=f(x)$ и $y_1=f(x)+b$ (графики этих функций получаются один из другого именно таким сдвигом) имеют в соответствующих точках одинаковые производные.

Можно заниматься и другой «математической игрой»: нарисовав от руки график производной, приблизительно построить график самой функции. При этом нужно произвольно задать одну точку $(x_0,\ y(x_0))$ графика и от нее вести кривую вверх или вниз (в соответствии со знаком производной).

В заключение остановимся на еще одном весьма важном обстоятельстве, тесно связанном, как мы увидим ниже, с обсуждаемым вопросом о связи наклона касательной к графику функции с производной (ранее мы этим обстоятельством намеренно пренебрегали). Заметим, что в физических задачах обычно x и $y=f(x)$ — размерные величины (например, x или t — это время, а $y=f(x)$ или $z=f(t)$ — расстояние), но тогда также и производная $\frac{dy}{dx}$ является размерной величиной. Ясно, что когда z измеряется в см, t — в с, производная $\frac{dz}{dt}=v(t)$ — это скорость, ее размерность — см/с. Если, скажем, y выражено в кг, а x — в месяцах (изучается зависимость $y=f(x)$ массы ребенка от его возраста), производная

$\frac{dy}{dx}\left(=\lim \frac{\Delta y}{\Delta x}\right)$, выражающая скорость увеличения массы, имеет размерность кг/мес; если y — сила тока, x — сопротивление, то производная $y'=\frac{dy}{dx}$ измеряется в А/Ом; если x — длина ребра куба, а y — его объем, то $\frac{dy}{dx}$ имеет размерность см^3/см$=$см^2, и т. д. Вообще, если аргумент x измеряется в единицах e_1, а функция $y=f(x)$ — в единицах e_2, то производная $\frac{dy}{dx}$ имеет размерность $\frac{e_2}{e_1}$.

Но величина $\operatorname{tg}\alpha$, как известно, безразмерна (она равна отношению двух отрезков, например двух катетов прямоугольного треугольника с острым углом α); отсюда уже ясно, что равенство $\frac{dy}{dx}=\operatorname{tg}\alpha$ не может иметь место (в левой части его стоит размерная величина, а в правой — безразмерная!). И лишь в тех редких случаях, когда обе переменные x и y имеют одинаковую размерность (скажем, когда изучается зависимость скорости парусной яхты от скорости ветра) или обе безразмерны (пример: функция $y=\sin x$, где x — радианная мера угла), соотношение $\frac{dy}{dx}=\operatorname{tg}\alpha$ можно считать осмысленным.

Как же исправить этот дефект наших рассуждений? Заметим, что ранее мы все время предполагали равенство масштабов по осям x и y, т. е. считали, что «единица измерения иксов» и «единица измерения игреков» изображаются на чертеже одинаковыми отрезками: это соглашение казалось нам настолько естественным, что мы даже не оговаривали его специально. Но в случае разной размерности переменных x и y (а именно этот случай надо считать общим!) сделанное предположение не только не естественно, но даже вообще не имеет смысла: нельзя же, в самом деле, всерьез полагать, что 1 с$=$1 см! Поэтому реально здесь очень часто пользуются р а з н ы м и масштабами по осям, вследствие чего соотношение $\frac{dy}{dx}=\operatorname{tg}\alpha$ нарушается.

Пусть, например, y есть путь, x — время и строится график $y(x)$ зависимости положения тела от времени x. По оси ординат будем откладывать y в масштабе: 1 м пути равен 1 см на чертеже. По оси абсцисс (x) будем откладывать время в масштабе: 1 с времени равна 1 см на чертеже. Тогда, действительно, значение скорости движения v, выраженное в метрах в секунду и равное производной $\frac{dy}{dx}$, совпадает со значением $\operatorname{tg}\alpha$, т. е. с тангенсом угла, образованного касательной к графику с осью x. Но если мы выберем другой масштаб для шкалы x, например 1 с$=l$ см на чертеже, то получим

$\operatorname{tg}\alpha=\frac{dy}{l\,dx}=\frac{1}{l}\frac{dy}{dx}$.

В частности, если $l=5$, мы имеем $\operatorname{tg}\alpha=\frac{1}{5}\frac{dy}{dx}=\frac{1}{5}v$.

В общем случае, если выбранные нами масштабы таковы, что одна единица x на чертеже изображается отрезком длины l_1 см, а одна единица y — отрезком длины l_2 см, то

$$\operatorname{tg}\alpha=\frac{l_2}{l_1}\frac{dy}{dx}. \qquad (2.5.3)$$

Масштабные множители l_1 и l_2 в формуле (3) исправляют положение — они делают это равенство правильным с точки зрения соображений размерности. Так, в примере с зависимостью массы ребенка от возраста множитель l_1 имеет размерность см/мес (он указывает, сколько сантиметров на графике приходится на 1 месяц возраста); аналогично l_2 имеет размерность см/кг (этот коэффициент указывает, сколько сантиметров на графике приходится на 1 кг массы). Таким образом, выражение $\frac{l_2}{l_1}\frac{dy}{dx}$ безразмерно: в формуле все размерности сокращаются, что и делает эту формулу осмысленной (и правильной).

Обо всем этом следует всегда помнить при сравнении производной функции и наклона графика этой функции (т. е. наклона касательной к графику в соответствующей точке).

Упражнения

2.5.1. Постройте график функции $y=x^2+1$ в пределах $-1{,}5 \leqslant x \leqslant 2{,}5$; проведите касательные к графику в точках $x=-1$; 0; 1; 2.

2.5.2. Постройте график функции $y=x^3-3x^2$ в пределах $-1 < x < 3{,}5$; проведите касательные к графику в точках $x=-1$; 0; 3. Укажите точки графика, касательная в которых горизонтальна.

2.5.3. Укажите на кривой $y=x^3-x+1$ такие точки, касательная в которых горизонтальна. Постройте кривую при $-2<x<2$. [Указание. Упр. 1—3 желательно выполнять на миллиметровой бумаге и в крупном масштабе.]

2.5.4. График функции $y(x)$ изображен на рис. 5; постройте (на глаз) график функции $y'(x)$.

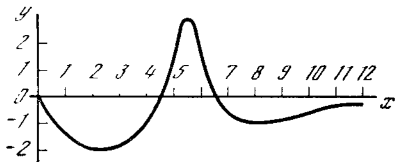


Рис. 2.5.5

2.5.5. График функции $y'(x)$ изображен на рис. 6. Постройте (на глаз) график функции $y(x)$, проходящий через точку (5, 0). Под каким углом последний график пересекает

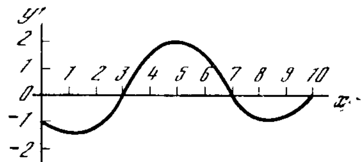


Рис. 2.5.6

ось ординат? Под каким углом он пересекает в точке $x=5$ ось абсцисс? [Указание. В упр. 4 и 5 следует сначала перерисовать изображенные на рис. 5 и 6 графики на чистый лист бумаги: требуемые графики целесообразно чертить на том же листе бумаги, например строго под заданным в условии задачи графиком.]

2.5.6. Составьте уравнения касательных к кубической параболе $y=x^3$ в точках: а) $x=0,5$; б) $x=1$. Найдите точки пересечения этих касательных с осями координат.

2.5.7. Укажите общее правило, позволяющее для кривых: а) $y=ax^2$; б) $y=bx^3$ — найти точки пересечения касательной к кривой в точке $(x_0,\ y(x_0))$ с осями координат.

§ 6. Рост и убывание функции. Максимумы и минимумы

Наблюдая ход линии, изображающей график некоторой функции $y=f(x)$ (скажем, зависимость $T(t)$ температуры T от времени t), мы без труда усмотрим на ней точки *роста* функции (вроде точки A на рис. 1) и точки ее *убывания* (например, точка B), точки *максимума* (наибольшего значения), где рост функции сменяется убыванием (точка C), и точки *минимума* (наименьшего значения), где, напротив, убывание сменяется ростом (точка D). Как же строго определить все эти понятия и как установить характер функции в данной точке $x=x_0$, не обращаясь к графику, вычерчивание которого зачастую вовсе не просто и который всегда ненадежен, ибо рисуется с неизбежными погрешностями (ошибками)?

Если не пользоваться производными, то ответ на поставленные вопросы нужно искать численно: например, найти температуру T в какой-то момент t; затем взять какой-то следующий момент t_1 и посмотреть, выросла температура или упала. Очевидно, что такой способ ненадежен: если $T(t_1)$ и больше, чем $T(t)$, то ведь не исключено, что в момент t температура падала, затем (вскоре после t, но еще до t_1) она достигла минимума, после этого стала расти, причем так, что к моменту t_1 оказалась выше, чем в момент t.

С помощью производных вопрос решается точно: надо найти производную $\frac{dy}{dx}$. Если величина $\frac{dy}{dx}=y'(x)$ при заданном x положительна, то $y(x)$ есть растущая функция: при увеличении x на малую величину Δx значение y изменится на малую величину $\Delta y \approx y'(x)\Delta x$ (как было выяснено раньше, чем меньше Δx, тем точнее это приближенное равенство). Мы рассматриваем случай $\Delta x>0$ (скажем, увеличение времени x). Если $y'(x)>0$ и $\Delta x>0$, то, конечно, и $\Delta y>0$, т. е. с течением времени величина y (например, температура) растет.

При этом численное значение (положительной) производной указывает, *как быстро* растет температура: если $T'(t)=10$, то вблизи от момента t температура растет в 10 раз быстрее,

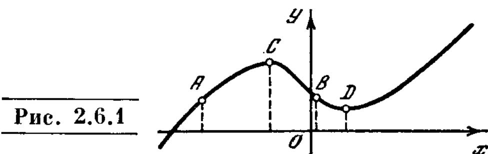


Рис. 2.6.1

чем время (например, на 10° за каждую секунду), а при $T'(t)=0,1$ она растет в 10 раз медленнее времени (ср., впрочем, со сказанным в конце § 5 о размерности производной и ее связи с выбором единиц измерения аргумента

и функции). Если же $y'(x) < 0$, то при $\Delta x > 0$ мы будем иметь $\Delta y < 0$: так, если $T'(x) < 0$, то температура $T(t+\Delta t)$ в момент времени $t+\Delta t$ будет ниже температуры $T(t)$ в данный момент. Таким образом, положительная производная указывает на то, что функция является *растущей*, а отрицательная — на то, что функция является *убывающей*, падающей.

Выражения «растущая функция» и «убывающая (падающая) функция» применяются не только к зависимостям величин от времени, но и к любой функции $y(x)$, аргументом которой служит произвольная (размерная или безразмерная) величина x; при этом *растущей* (*убывающей*) называется такая функция, которая увеличивается (соответственно уменьшается) с ростом независимой переменной x. Производная $\frac{dy}{dx}$ характеризует *скорость роста*: она равна отношению изменения y к отвечающему ему (небольшому) изменению x. Отрицательная скорость роста означает падение, уменьшение y при увеличении x, и если $\frac{dy}{dx} < 0$, то $\left|\frac{dy}{dx}\right| = -\frac{dy}{dx}$ есть скорость убывания функции.

Выражение «величина y имеет большую отрицательную производную по x» означает, что y быстро падает с увеличением x, а положительная производная $\frac{dy}{dx}$ означает, что y с ростом x увеличивается. Таким образом, производная переменной y указывает *тенденции*, характерные для изменения y, позволяет судить о том, что можно ожидать при дальнейшем изменении аргумента — в этом заключается значение величины y'.

Физики и математики (в особенности будущие физики и математики, только недавно узнавшие, что такое производная) часто применяют это понятие и в повседневной жизни: «производная от моего настроения по времени положительна» — вместо «мое настроение улучшается».

Решите задачу-шутку: какой знак имеет производная от настроения по расстоянию до кресла зубного врача? При уменьшении расстояния до кресла — по мере приближения к креслу — наше настроение ухудшается, можно также сказать «падает», становится «отрицательным», значит, производная здесь положительна.

Может быть, писатели и будут сетовать на засорение языка, но на самом деле такое вольное шуточное употребление математических понятий — хорошая тренировка для будущих серьезных применений математики.

Существуют функции, производная которых сохраняет один и тот же знак при любых значениях переменной. Таким свойством обладает, например, линейная функция $y=kx+b$: ведь здесь $\frac{dy}{dx} = k$, т. е. постоянная величина. Позже мы увидим, что также и производная показательной функции $y=a^x$ при любых x имеет один и тот же знак (хотя и не постоянна по величине).

Представим себе функцию $y(x)$, производная которой $y'(x)$ положительна при $x < x_0$ и отрицательна при $x > x_0$ (кратко: $y'(x) > 0$ при $x < x_0$; $y'(x) < 0$ при $x > x_0$).

Что можно сказать о такой функции? Начнем со значений $x < x_0$. При увеличении x до x_0 величина y будет расти; однако при дальнейшем увеличении x производная становится отрицательной и величина y уменьшается. Отсюда вывод: при $x=x_0$ функция $y(x)$ имеет максимум.

Рассмотрим противоположный случай: $y'(x) < 0$ при $x < x_0$; $y'(x) > 0$ при $x > x_0$. Рассуждая так же, как и выше, мы придем к выводу, что в этом случае при $x=x_0$ функция $y(x)$ имеет минимум.

Если функция $y(x)$ задана формулой, которой соответствует плавная кривая, то и $y'(x)$ задается формулой и плавно изменяется при изменении x. Но в этом случае разные знаки $y'(x)$ при $x < x_0$ и при $x > x_0$ означают, что при $x=x_0$ производная *обращается в нуль*:

$$y'(x_0) = \frac{dy(x_0)}{dx} = 0. \qquad (2.6.1)$$

Таким образом, как мы уже отмечали в предыдущем параграфе, *приравнивая нулю производную, можно найти те значения независимой переменной, при которых функция имеет максимум или минимум* [1]. Теперь мы можем уточ-

[1] Точнее, конечно, значения, при которых функция может иметь (а может и не

нить общую теорему Ферма (см. § 5): в точках, где функция достигает *максимума*, производная переходит *от положительных значений к отрицательным*, а в точках *минимума* она переходит *от отрицательных значений к положительным*.

Приведем несколько примеров. Прежде всего обратимся к рассмотренной в § 1.1 функции $y=3x^3-x^2-x$ (см. таблицу в § 1.1). Судя по таблице значений функции, можно думать, что функция является растущей при всех значениях x, так как каждое увеличе-

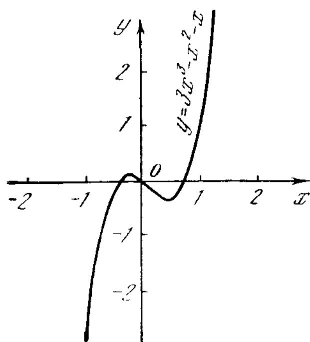


Рис. 2.6.2

ние x на единицу вызывает увеличение y.

Вычислим, однако, производную нашей функции:

$$y'(x)=9x^2-2x-1.$$

Взяв $x=0$, получим $y'(0)=-1<0$. Значит, при $x=0$ функция у б ы в а ю щ а я; это опровергает полученное из рассмотрения таблицы предположение, что наша функция растет везде.

Приравняем $y'(x)$ нулю. Уравнение

$$y'(x)=9x^2-2x-1=0$$

имеет два корня:

$$x_1\approx-0{,}24,\quad x_2\approx0{,}46.$$

Составим подробную таблицу, включая в число рассматриваемых значений x и найденные точки обращения в нуль производной:

x	—3	—2	—1
y	—87	—26	—3
x	—0,3	—0,24	—0,18
y	0,129	0,140	0,131
x	0	0,40	0,46
y	0	—0,372	—0,381
x	0,52	1	2
y	—0,370	1	18

Мы видим, что на участке от $x\approx-0{,}24$, до $x\approx0{,}46$ функция y уменьшается от 0,14 до —0,38. Сравнение значения $y(-0{,}24)$ с «соседними» $y(-0{,}30)$ и $y(-0{,}18)$ подтверждает, что при $x\approx-0{,}24$ функция $y(x)$ достигает (локального, или «местного») максимума, ибо соседние значения y меньше. График функции $y=3x^3-x^2-x$ изображен на рис. 2.

Этот пример снова показывает, что максимум не означает наибольшего из всех вообще возможных значений y. Ведь у нас в точке максимума $y(-0{,}24)\approx0{,}14$, а при $x=1$ имеем $y=1$, при $x=2$ получаем $y=18$, а при $x=10$ — даже $y=2890$; вообще при неограниченном увеличении аргумента x функция y также неограниченно и быстро растет. Так что же выделяет точку максимума $x_{max}\approx-0{,}24$, $y_{max}\approx0{,}14$ из множества всех значений аргумента?

Отличие этой точки от других состоит в том, что при б л и з к и х (так сказать, соседних) с x_{max} значениях x, как бо́льших x_{max}, так и меньших x_{max}, величина y меньше, чем y_{max} $(=y(x_{max}))$. Эта особенность x_{max} наглядно просматривается в приведенной выше таблице (ср., например, $y(-0{,}3)$, $y(-0{,}24)$ и $y(-0{,}18)$). Те же соображения относятся и к точке минимума $x_{min}\approx0{,}46$, $y_{min}\approx-0{,}381$: хотя y при больших по абсолютной величине отрицательных x неограниченно уменьшается и становится меньше y_{min} (и вообще может быть сделан меньше любого отрицательного числа!), но величина x_{min} отличается тем, что значение y_{min}

иметь!) максимум или минимум. Так, функция $y=x^3$ имеет производную $y'=3x^2$, обращающуюся в 0 при $x=0$; однако в этой точке функция (см. ее график на рис. 1.5.1, *а*) не имеет ни максимума, ни минимума. (О других исключениях из нашего правила, касающихся случаев негладких кривых, будет сказано в § 7.2.)

($=y\,(x_{\min})$) меньше всех y, отвечающих «соседним» с $x_{\min}$ значениям аргумента (значениям x, достаточно близким к $x_{\min}$). Условие равенства нулю производной дает возможность найти именно такие — *локальные*, или *местные*, — максимумы и минимумы.

В качестве второго примера рассмотрим фигурирующую в § 1.5 функцию $y=x^3-x$. Производная этой функции $y'=3x^2-1$; поэтому $y'(x)=0$ при $x=\pm 1/\sqrt{3}$. Проследив за изменением знака функции $f'(x)=3x^2-1$ в окрестностях точек $x=1/\sqrt{3}\approx 0{,}577$ и $x=-1/\sqrt{3}\approx -0{,}577$, мы без труда установим, что при $x=-1/\sqrt{3}$ функция y имеет (локальный) максимум, а при $x=1/\sqrt{3}$ — (локальный) минимум (см. рис. 1.5.3).

Теперь мы можем полнее раскрыть причины различия между изображенными на рис. 1.5.2 и 1.5.3 графиками функций $y=x^3+x$ и $y=x^3-x$. Рассмотрим общее уравнение

$$y=x^3+cx, \qquad (2.6.2)$$

где c — произвольное число. В этом случае $y'=3x^2+c$. Ясно, что при $c>0$ величина y' всюду положительна, значит, y растет при всех x — ни максимумов, ни минимумов функция не имеет. При $c<0$, напротив, $y'=0$ при $x=\pm\sqrt{-\frac{c}{3}}=\pm x_0$, причем в интервалах значений x от $-\infty$ до $-x_0$ производная y' положительна (функция возрастает); при $-x_0<x<x_0$ производная отрицательна (функция убывает); при $x_0<x<\infty$ производная снова положительна (функция снова растет; здесь символы $-\infty$ и ∞, как всегда, обозначают очень большие отрицательное и положительное числа). Таким образом, при $x=-x_0$ наша функция имеет максимум, а при $x=x_0$ — минимум. При уменьшении c по абсолютной величине эти максимум и минимум сближаются: при $c=0$ точки $\pm x_0=\pm\sqrt{-\frac{c}{3}}$ сливаются в единственную «критическую точку» $x=0$ — точку нулевой производной функции y. В этом случае (для кривой $y=x^3$) соответствующая точка уже не является ни максимумом, ни минимумом функции: ведь из рассмотрения «соседних» кривых, отвечающих малым по абсолютной величине отрицательным c, вроде бы следует, что наша точка должна являться одновременно и максимумом и минимумом, что явно невозможно; ожидать же здесь только максимума (не минимума) или только минимума (не максимума) у нас нет никаких оснований. Все эти особенности кривых (2) изображены на рис. 3 (у кривых проставлены отвечающие им значения c). Из этого рисунка видно, что существует два принципиально разных типа кривых, отвечающих случаям $c<0$ и $c>0$; «промежуточной» между этими

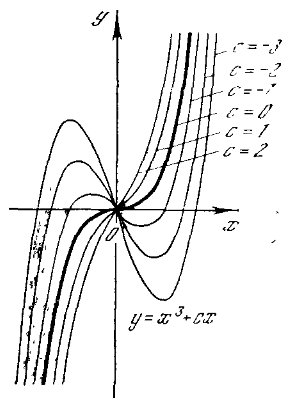


Рис. 2.6.3

типами является «предельная» кривая $y=x^3$, для которой $c=0$.

Возвратимся снова к упр. 2.1, посвященному обсуждению вопроса о теплоемкости алмаза. В этом упражнении фигурировала следующая экспериментально полученная зависимость от температуры T количества $Q=Q\,(T)$ тепла (в Дж), необходимого для нагревания 1 кг алмаза от 0 до T° С:

$$Q\,(T)=0{,}3965T+2{,}081\cdot 10^{-3}\,T^2-5{,}024\cdot 10^{-7}\,T^3,$$

откуда вытекает, что теплоемкость $c=c\,(T)$ алмаза (в Дж/(кг·°С)) выражается формулой

$$c\,(T)=0{,}3965+4{,}162\cdot 10^{-3}\,T-15{,}072\cdot 10^{-7}\,T^2$$

(см. решение упр. 2.1). Для того чтобы проанализировать характер изменения теплоемкости алмаза в рассматриваемом диапазоне температур (том, при котором может использоваться экспериментальная формула упр. 2.1), нам придется составить производную $c'\,(T)$ теплоемкости c:

$$c'\,(T)=4{,}162\cdot 10^{-3}-30{,}144\cdot 10^{-7}\,T.$$

Ясно, что величина $c'\,(T)$ положительна при $T<(4{,}162/30{,}144)\cdot 10^4=T_0\approx 1380^\circ$ С; равна нулю при $T=T_0$, отрицательна при $T>T_0$.

Но ведь сама эмпирическая формула, как указывалось выше, справедлива только до 700—800° С. Формально из выписанных фор-

мул следует, что при $T \approx 1380°C$ обращается в нуль величина $c'(T)$, при $T \approx 2850°C$ обращается в нуль сама теплоемкость $c(T)$ и, наконец, при $T \approx 4320°C$ обращается в нуль количество тепла $Q(T)$. Все эти результаты, разумеется, бессмысленны. Они только лишний раз доказывают недопустимость использования эмпирических формул за пределами области их применимости!

Определение максимумов и минимумов арифметическим путем — вычислением и сравнением значений функции при различных значениях аргумента — является во много раз более трудоемким и менее точным. Правда, если в руках у Вас есть компьютер, то «арифметический эксперимент», позволяющий обнаружить максимум или минимум путем разумного сравнения значений функции в разных точках (ведь компьютером тоже, конечно, надо пользоваться умело, а не как попало), значительно облегчается, — но даже и в этом случае использование понятия производной много проясняет в существе дела, делает выкладки и подсчеты более «прозрачными» и простыми. Высшая математика является отнюдь не только замечательным идейным достижением. Практические, конкретные вычислительные задачи также гораздо легче решаются методами высшей математики.

Упражнения

2.6.1. Найдите значения x, при которых достигается максимум или минимум следующих функций. В каждом случае выясните, имеем ли мы дело с минимумом или с максимумом. В функциях, в которые входят постоянные, специально рассмотрите случаи разных (в частности, разных по знаку) значений этих постоянных: а) $y = ax^2$; б) $y = x + \frac{1}{x}$; в) $y = ax + \frac{b}{x}$; г) $y = x^3 - 3x + 100$; д) $y = x^3 + px^2 + qx + r$; е) $y = x^4 + ax^2 + b$; ж) $y = ax^2 + \frac{b}{x^2}$.

2.6.2. Объем $v = v(t)$ 1 г воды (в см^3) выражается следующей (приближенной) эмпирической формулой: $v(t) = 1 + 8{,}38 \cdot 10^{-6}(t-4)^2$. При какой температуре этот объем будет минимальным?

2.6.3. Ответьте на вопрос упр. 2 исходя из следующих (предложенных разными естествоиспытателями) уточнений формулы для $v(t)$: а) $v(t) = 1 - 61{,}045 \cdot 10^{-6}t + 77{,}183 \cdot 10^{-7}t^2 - 37{,}34 \cdot 10^{-9}t^3$; б) $v(t) = 1 - 57{,}577 \cdot 10^{-6}t + 75{,}601 \cdot 10^{-7}t^2 - 35{,}07 \cdot 10^{-9}t^3$.

§ 7. Вторая производная функции. Выпуклость и вогнутость кривой. Точки перегиба

Вернемся снова к рассмотрению «критических» (подозрительных на максимум и минимум) точек $x = x_0$ кривой $y = y(x)$, т. е. таких, что

$$y'(x_0) = 0. \qquad (2.7.1)$$

Нас интересует вопрос о том, как в случае выполнения (1) отличить максимум от минимума: ведь (1) выполняется и в точках (локального) максимума и в точках (локального) минимума.

Мы уже знаем, что разница заключается в знаке производной $y'(x)$ при $x < x_0$ и при $x > x_0$. Но как определить знак $y'(x)$ при x, близком к x_0, не вычисляя непосредственно y' для других значений x? Обратимся к случаю максимума функции $y(x)$, когда $y'(x) > 0$ при $x < x_0$ и $y'(x) < 0$ при $x > x_0$. Мы видим, что в этом случае производная $y'(x)$ сама представляет собой убывающую функцию: по мере роста x производная, которая сначала была положительной (при $x < x_0$), обращается в нуль (при $x = x_0$) и затем становится отрицательной (при $x > x_0$). Но мы уже знаем, как отличить убывающую функцию от возрастающей: производная убывающей функции *отрицательна*. Следовательно, при том значении $x = x_0$, при котором y имеет *максимум*, $y'(x_0) = 0$, а производная от производной *отрицательна*. Такая величина — производная от производной, которую по общим правилам можно записать «трехэтажной» дробью

$$\frac{dy'}{dx} = \frac{d\left(\frac{dy}{dx}\right)}{dx},$$

имеет свое название — *вторая производная*. Ее обозначают также $y''(x)$, или $\frac{d^2y}{dx^2}$ (читается: «де два игрек по де икс квадрат»).

Ясно, что если $z = z(t)$ — зависимость пути от времени (см. § 1 наст. гл.), то $z'(t) = v$ — скорость движения, а $z''(t) = \frac{dv}{dt}$ — «скорость изменения скорости», т. е. *ускорение* (см. также гл. 9). Если y имеет размерность пути (в см), а x — это время (в с), то $\frac{d^2y}{dx^2}$ имеет

размерность ускорения (см/с^2); если y измеряется в каких-то единицах e_1, а x — в единицах e_2, то производная $\frac{d^2y}{dx^2}$ имеет размерность $e_1/(e_2)^2$, как это и указывается самой записью $\frac{d^2y}{(dx)^2}$; эту размерность второй производной полезно не забывать при рассмотрении физических задач, в которых фигурирует величина $y''(x)$.

Именно тот смысл, который имеет вторая производная (ускорение) в нашем основном примере путь—время, делает понятие второй производной особенно важным для физики: ведь в силу второго закона Ньютона именно ускорение является главной характеристикой движения. Подробнее мы еще остановимся на этом в гл. 9.

Обратимся теперь снова к рассмотренным выше примерам. Если $y(x) = 3x^3 - x^2 - x$, то $y'(x) = 9x^2 - 2x - 1$, и, значит,

$$y''(x) = (y'(x))' = 18x - 2.$$

При $x \approx -0{,}24$ имеем $y' = 0$, а $y'' \approx -6{,}3 < 0$ — и, действительно, точка $x \approx -0{,}24$, $y \approx 0{,}14$, есть точка максимума. При $x \approx 0{,}46$ имеем: $y' = 0$, а $y'' \approx 6{,}3 > 0$, т. е. при $x \approx 0{,}46$, $y \approx -0{,}38$ функция имеет минимум.

Аналогично, если $y = x^3 - x$, то $y' = 3x^2 - 1$ и $y'' = 6x$. Поэтому при $x = -1/\sqrt{3} \approx -0{,}577$ вторая производная y'' отрицательна, а при $x = +1/\sqrt{3}$ она положительна: в первой точке функция y имеет максимум, а во второй — минимум. (Впрочем, здесь этот результат можно предвидеть и геометрически: так как при $x = -1$ и при $x = 0$ функция $y = x^3 - x = -(x - x^3)$ обращается в нуль, а в промежуточных точках она положительна, то при $x \approx -0{,}577$ мы можем иметь только максимум, но никак не минимум.) Точно так же и в примере с теплоемкостью алмаза (см. § 6) вторая производная $c''(T) \approx -0{,}003$ всегда отрицательна, что обеспечило бы максимум $c(T)$, если бы наши формулы были применимы в точке T_0, где $c'(T_0) = 0$ (реально теплоемкость алмаза, как и других твердых тел, все время растет с ростом T, стремясь при высокой температуре к постоянной величине). Итак,

1) если при определенном значении x функция $y(x)$ такова, что

$$y'(x) = 0, \quad y''(x) < 0, \tag{2.7.2а}$$

то в этой точке функция имеет максимум;

2) если при некотором x функция удовлетворяет условиям

$$y'(x) = 0, \quad y''(x) > 0, \tag{2.7.2б}$$

то в этой точке функция имеет минимум. (Рекомендуем теперь вернуться к приведенным в конце § 6 упражнениям и применить к ним «критерий второй производной», позволяющий отличить максимум от минимума.)

Мы видели, что знак производной $y = f(x)$ имеет простой геометрический

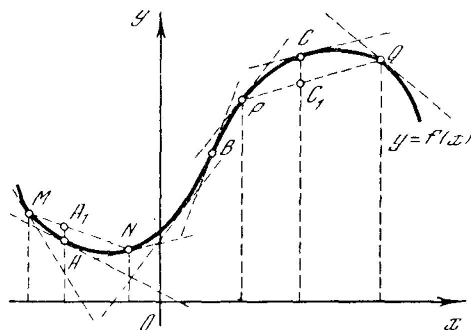


Рис. 2.7.1

смысл: положительность производной означает, что наша функция является возрастающей, т. е. что график ее идет вверх, если следовать по нему в направлении слева направо; при $y'(x) < 0$ функция убывает, т. е. график ее идет вниз. Знак второй производной $y''(x)$ также имеет достаточно прозрачный геометрический смысл. Факт положительности второй производной $y''(x) > 0$ означает, что первая производная $y'(x)$ растет, другими словами, растет угол α (где $\operatorname{tg} \alpha = k = y'(x)$), образуемый касательной к графику функции с осью абсцисс. Но в таком случае при движении слева направо по графику функции касательная к графику поворачивается в направлении, обратном направлению вращения часовой стрелки (см. точку A на рис. 1). А это, очевидно, означает, что наша кривая расположена *выпуклостью вниз* (ср. § 1.4); иногда еще в этом случае говорят, что кривая *вогнута вверх*. Аналогично неравенство $y''(x) < 0$ означает, что первая производная $y'(x)$ функции $y = f(x)$ убывает, а значит, убывает и угол α. Таким образом, в точках, где $y''(x) < 0$, каса-

тельная к графику функции при движении (слева направо) по кривой вращается в направлении вращения часовой стрелки, т. е. кривая *выпукла*, обращена *выпуклостью вверх* (см. точку C на рис. 1). Точки, в которых $y''(x)=0$, вообще говоря, можно охарактеризовать как точки изменения знака второй производной, т. е. изменения направления выпуклости кривой [1]; в этих точках

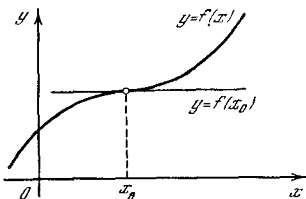


Рис. 2.7.2

касательная к графику функции y переходит с одной стороны графика на другую сторону, т. е. здесь касательная *пересекает* график. Подобные точки, в которых касательная к кривой пересекает саму кривую (см. точку B на рис. 1), принято называть *точками перегиба* [2].

В частности, когда график функции $y=f(x)$ пересекает горизонтальная касательная, проведенная в точке $x=x_0$ (рис. 2), точка x_0 не является ни максимумом, ни минимумом функции, а между тем производная в ней равна нулю (касательная горизонтальна!). Таким образом, мы видим, что случаи, когда равенство (1) «работает вхолостую» — не указывает ни максимума, ни минимума функции, связаны с тем, что в рассматриваемой точке выполнено не только условие (1), но и условие

$$y''(x_0)=0, \qquad (2.7.3)$$

характеризующее точку перегиба. (Ср., впрочем, со сказанным в сноске 1 о точке $x=0$ функции $y=x^4$, в которой выполняются оба условия (1) и (3), а функция тем не менее и м е е т минимум.)

Ниже (см. § 7.3) мы еще встретимся снова с понятием выпуклой функции.

Упражнения

2.7.1. Найдите вторые производные функций $y=x^2$; $y=x^3$; $y=x^4$; $y=ax^2+bx+c$.

2.7.2. Найдите ускорение движения точки по прямой, задаваемого следующей зависимостью пути от времени: $z=at^2+bt+c$. Что можно сказать об этом ускорении?

2.7.3. Укажите участки выпуклости и вогнутости графика функции: а) $y=x^3$; б) $y=x^3+px^2+qx+r$; в) $y=x+\frac{1}{x}$.

[1] Редкие исключения типа начала координат для кривой $y=x^4$ (см. рис. 1.5.1, *б*; здесь $y'=4x^3$, $y''=4\cdot 3x^2=12x^2$, так что $y''=0$ при $x=0$, а между тем линия в этой точке направления выпуклости не меняет) относятся исключительно к точкам, в которых одновременно обращаются в нуль и вторая, и третья производные функции (или вторая производная имеет излом либо разрыв). (Третья производная — это производная от второй производной; она записывается как y''', или $\frac{d^3y}{dx^3}$, и имеет размерность $e_1/(e_2)^3$, где e_1 и e_2 — единицы измерения величин y и x. Ниже нам еще предстоит встретиться с третьей и последующими производными функции.)

[2] Заметьте, в точке перегиба касательная пересекает кривую «плавно», с «соприкосновением», а не «грубо», как, скажем, ось y пересекает параболу $y=x^2$.

Глава 3

ЧТО ТАКОЕ ИНТЕГРАЛ

§ 1. Определение пути по скорости движения и определение площади, ограниченной кривой

Задача об определении мгновенной скорости движения $v(t)$ по заданной зависимости $z = z(t)$ положения z тела от времени t привела нас к понятию производной:

$$v(t) = \frac{dz}{dt}.$$

Обратная задача заключается в определении положения $z = z(t)$ тела (т. е. пути, пройденного телом за данный отрезок t времени), если мгновенная скорость $v(t)$ задана как функция времени. Эта задача приводит ко второму важнейшему понятию высшей математики — понятию **интеграла**.

Условимся об обозначениях, которые будут нам удобны. Мы будем рассматривать путь, пройденный за время от момента t_1 до момента t_2. Чтобы не писать индексов, обозначим начало рассматриваемого промежутка времени буквой a, конец этого промежутка — буквой b; таким образом, у нас $t_1 = a$, $t_2 = b$. Пройденный (за время от a до b) путь обозначим через $z(a, b)$. Запомним, что запись $z(a, b)$, где в скобках под знаком функции z стоят две величины a и b, обозначает длину пути, пройденного между моментами $t = a$ и $t = b$, тогда как функция $z(t)$, где в скобках указано одно число t, задает положение (координату) тела в заданный момент времени t. Очевидно, между величинами $z(a, b)$ и $z(t)$ существует простая связь:

$$z(a, b) = z(b) - z(a),$$

т. е. $z(b) = z(a) + z(a, b)$ (3.1.1)

(мы здесь считаем, что $b > a$), — путь $z(a, b)$, пройденный за время от a до b, равен разности $z(b) - z(a)$ координат тела в начале и в конце рассматриваемого интервала времени [1].

[1] Для простоты мы в этих наглядных рассмотрениях ограничиваемся случаем движения тела в одном направлении (хотя, возможно, с переменной скоростью).

Обратимся теперь к вычислению величины $z(a, b)$. В простейшем случае, когда скорость v постоянна:

$$v(t) = \text{const} = v_0, \qquad (3.1.2)$$

пройденный путь равен просто произведению времени движения на скорость:

$$z(a, b) = (b - a)\, v_0. \qquad (3.1.3)$$

Воспользуемся графиком зависимости скорости от времени, при постоянной скорости представляющим собой горизонтальную прямую (рис. 1). Пройденный путь, очевидно, равен заштрихованной площади, поскольку площадь

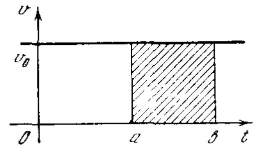


Рис. 3.1.1

прямоугольника равна произведению основания $(b - a)$ на высоту (v_0). А как быть в общем случае, когда скорость движения не постоянна?

Рассмотрим подробно один численный пример. Пусть скорость движения задана формулой [2] $v = t^2$. Найдем путь за время от $t = a = 1$ до $t = b = 2$.

Разобьем весь промежуток времени от a до b на десять частей и составим таблицу скоростей:

t	1,0	1,1	1,2	1,3	1,4	1,5
v	1,0	1,21	1,44	1,69	1,96	2,25

t	1,6	1,7	1,8	1,9	2,0
v	2,56	2,89	3,24	3,61	4,0

В чем трудность вычисления пути при скорости $v(t)$, заданной известной нам формулой? Очевидно, все дело в том, что скорость переменна (для по-

[2] Если скорость v выражена в см/с, а время t — в с, то, чтобы соблюдались требования размерности, надо писать $v = kt^2$, где коэффициент k имеет размерность см/с³. Здесь мы считаем, что $k = 1$ см/с³.

стоянной скорости задача элементарна). В рассматриваемом случае скорость в промежутке времени от $t=1$ до $t=2$ меняется в 4 раза. Однако после того как этот промежуток разбит на 10 частей, в каждом маленьком промежутке длительностью в 0,1 с скорость меняется гораздо менее резко — всего на 10—20%. Следовательно, в каждом таком промежутке времени скорость приближенно можно считать постоянной, что позволяет рассчитать проходимый за этот малый промежуток времени путь по простой формуле: скорость,

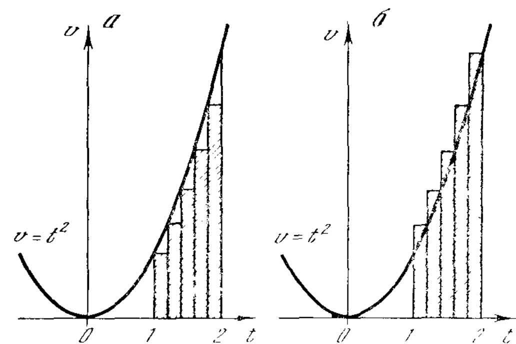


Рис. 3.1.2

умноженная на время. В дальнейшем малые промежутки времени, на которые мы разбиваем большой промежуток времени от $t=a$ до $t=b$, будем обозначать символом Δt («приращение времени»); для каждого такого промежутка времени Δt мы будем считать путь просто пропорциональным длительности Δt промежутка, полагая скорость на всем его протяжении одинаковой.

Для вычисления пути в каждом промежутке Δt, составляющем 0,1 с, используем скорость в н а ч а л е этого промежутка: 1 см/с в Δt от 1 до 1,1 с; 1,21 см/с в Δt от 1,1 до 1,2 с и т. д.; наконец, 3,61 см/с в последнем Δt от 1,9 до 2,0 с. Полный путь, пройденный за время от $t=1$ до $t=2$, при этом способе подсчета окажется равным

$$z(1,\ 2) \approx 0{,}1 + 0{,}121 + 0{,}144 + \ldots$$
$$\ldots + 0{,}361 = 2{,}185 \text{ см.} \qquad (3.1.4)$$

Очевидно, что при таком расчете мы *приуменьшили* пройденный путь: скорость в данном примере с течением времени растет, поэтому начальная скорость для каждого промежутка меньше средней скорости в этом промежутке. Каждое из десяти слагаемых, на которые разбит весь путь, несколько занижено; следовательно, занижен и весь результат.

Теперь подсчитаем путь по-другому, а именно в каждом промежутке Δt будем брать значение скорости в к о н ц е промежутка. Для первого Δt от 1 до 1,1 с эта скорость равна 1,21 см/с, для последнего Δt от 1,9 до 2 с скорость равна 4 см/с и т. д. Так мы получим следующую оценку всего пройденного пути:

$$z(1,\ 2) \approx 0{,}121 + 0{,}144 + \ldots$$
$$\ldots + 0{,}400 = 2{,}485 \text{ см.} \qquad (3.1.5)$$

Такой расчет, очевидно, дает *преувеличенное* значение $z(1,\ 2)$. Значит, истинное значение лежит в пределах между 2,185 и 2,485 см. Различие между 2,185 и 2,485 составляет около 15%. Округляя границы для z, получаем

$$2{,}18 < z(1{,}2) < 2{,}49.$$

Проделанный расчет можно пояснить графически. Построим график (рис. 2), на котором по оси абсцисс отложено время, а по оси ординат — скорость. (Для того чтобы ступеньки были хорошо видны, на рис. 2 рассматриваемый промежуток времени разбит на пять частей, а не на десять, как в тексте.) Каждое слагаемое в первой сумме (4) представляет собой площадь узкого прямоугольника, основанием которого является соответствующий интервал Δt, а высотой — скорость в начале этого интервала; сумма же всех слагаемых представляет собой (заштрихованную на рис. 2, *а*) площадь под ломаной (ступенчатой) линией. Вторая сумма (5), в которой для каждого интервала Δt времени скорость полагалась такой, как в конце этого интервала, равна площади, заштрихованной на рис. 2, *б*.

А как можно точнее подсчитать путь, пройденный за данное время (в нашем примере за время от $t=a=1$ с до $t=b=2$ с)? Ясно, что различие между «нижней» и «верхней» оценками пути (в нашем случае между 2,18 и 2,49) зависит от быстроты изменения скорости в каждом из малых интервалов Δt. Поэтому для того чтобы это различие стало меньше, надо просто позаботиться о том, чтобы скорость в пределах рассматри-

ваемых интервалов меньше менялась; но так как скорость от нас не зависит, то нам остается только уменьшать сами эти интервалы. Впрочем, этот путь нас полностью устраивает, поскольку требуемый результат он обеспечивает: если сделать все интервалы меньше (а общее их количество — соответственно больше), то полученные по описанному выше способу две оценки для пути $z(a, b)$ сблизятся между собой.

Так, например, если разбить тот же интервал от 1 до 2 с не на 10, а на 20 промежутков длительностью по 0,05 с каждый, то расчет, аналогичный проведенному выше, дает при условии учета в каждом промежутке Δt начальных скоростей

$$z(1, 2) \approx 0{,}05 + 0{,}05 \cdot 1{,}1025 + \ldots$$
$$\ldots + 0{,}05 \cdot 3{,}8025 = 2{,}25875. \quad (3.1.4а)$$

Если же мы заменим начальные скорости конечными, то получим

$$z(1, 2) \approx 0{,}05 \cdot 1{,}1025 + 0{,}05 \cdot 1{,}21 + \ldots$$
$$\ldots + 0{,}05 \cdot 4 = 2{,}40875. \quad (3.1.5а)$$

Различие между 2,25875 и 2,40875 составляет уже всего лишь около 7%. Пределы, в которых заключено $z(1, 2)$, сузились. Округленно:

$$2{,}25 < z(1, 2) < 2{,}41.$$

При уменьшении Δt результат приближается к *истинному* значению пути, которое будет вычислено в дальнейшем и окажется равным

$$z(1, 2) = 2^1/_3 \approx 2{,}333.$$

Ведь чем меньше Δt, тем меньше и различие в каждом малом промежутке Δt между начальной и конечной скоростями; следовательно, при уменьшении Δt уменьшается относительная ошибка в каждом слагаемом. Именно потому, что речь здесь идет об *относительной* ошибке, мы можем быть уверены, что хотя число слагаемых (т. е. количество ошибок) у нас увеличивается, все равно точность оценки общей суммы путей, т. е. точность оценки величины $z(1, 2)$, наверняка растет.

Для иллюстрации этого на рис. 3 изображены построения, связанные с оценкой пути $z(1, 2)$ по заданной скорости $v = t^2$ при $\Delta t = 0{,}5$; 0,2 и 0,1.

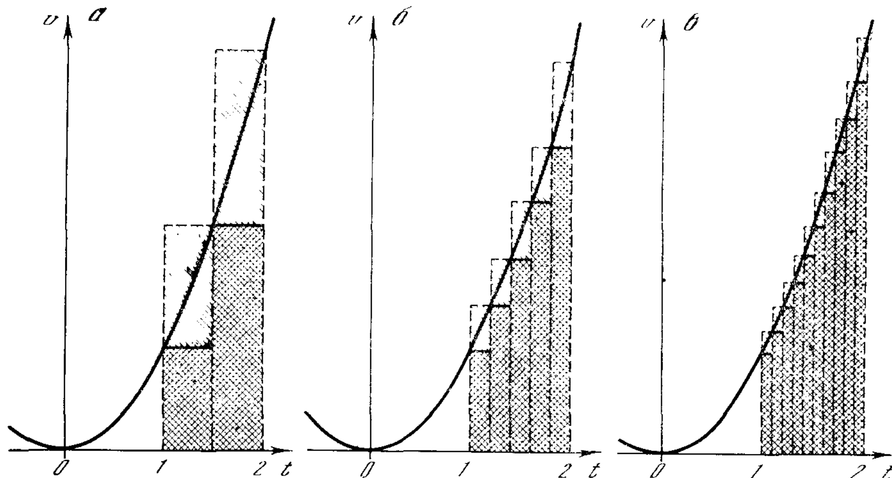


Рис. 3.1.3

Геометрически очевидно, что при увеличении числа промежутков Δt и уменьшении длины каждого промежутка уменьшаются размеры каждой ступеньки на рис. 3, поэтому ступенчатая линия становится все ближе к кривой $v(t)$.

Суммируя все наши рассмотрения, мы приходим к выводу, что при произвольной зависимости $v = v(t)$ мгновенной скорости v от времени t путь, пройденный за время от $t = a$ до $t = b$, равен *площади* фигуры, ограниченной линией $v = v(t)$, вертикалями $t = a$ и $t = b$ и осью абсцисс (ось t; рис. 4).

Этот вывод дает способ практического вычисления пути: можно построить график на миллиметровке и определить заштрихованную площадь либо подсчетом клеточек, либо, например, вырезав эту площадь из картона или бумаги, взвесив вырезанный листочек и сравнив его массу с массой вырезанного из того же материала прямоугольника или квадрата известной площади [3].

[3] Разумеется, при наличии (микро)компьютера метод, связанный с рассмотрением

Подобные методы уместны и вполне оправданы, если скорость известна не точно, т. е. задана не формулой, а таблицей или графиком, полученными экспериментальным путем. Мы не остановимся на этих приближенных способах: наше внимание будет сосредоточено на том, как выразить пройденный путь **формулой**, когда и зависимость скорости от времени также задана **формулой**.

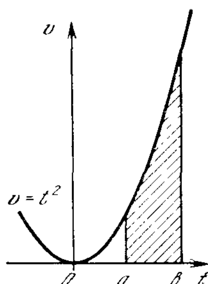


Рис. 3.1.4

Укажем еще, что в отдельных случаях, например когда скорость на интервале от 0 до 1 с выражается формулой $v(t)=\frac{1}{1-t}$ и при приближении к значению $t=1$ с неограниченно растет, путь может вовсе не существовать (становиться бесконечным); такие случаи также пока будут оставлены без внимания (ср., впрочем, § 5.2).

Можно указать и более точный (по сравнению с описанным выше) численный метод определения пути по скорости. Условимся в каждом промежутке Δt времени считать скорость постоянной и равной *среднему арифметическому* (т. е. полусумме) начальной и конечной скоростей в этом промежутке. При таком способе при разбиении на десять промежутков скорость в первом промежутке от 1 до 1,1 с принимается равной $\frac{1+1{,}21}{2}=1{,}105$ см/с (см. таблицу в начале этого параграфа) и путь за этот промежуток времени считается равным 0,1105 см; путь за второй промежуток подсчитывается как $0{,}1\cdot\frac{1{,}21+1{,}44}{2}=0{,}1325$ см и т. д. Складывая все полученные таким способом пути, найдем путь, пройденный за все время от $a=1$ с до $b=2$ с; в нашем случае он будет равен

$z(1,2)=0{,}1105+0{,}1325+\ldots=2{,}335$ см.

При разбиении на 20 промежутков получим таким же способом, считая среднюю скорость равной полусумме скоростей (в начале и в конце), $z(1,\ 2)=2{,}33375$ см.

таблицы значений функции и составлением сумм типа (4), (5) (или (4а), (5а)), удобнее и точнее примитивного взвешивания фигуры, вырезанной из бумаги или из картона.

Эти значения гораздо ближе к истинной величине 2,33333, чем вычисленные по начальному или конечному значению скорости при том же числе промежутков: при десяти промежутках мы получаем ошибку в 0,07% вместо 15% ранее, а при 20 промежутках — ошибку в 0,02% вместо 7%.

Новый способ нахождения пути можно также наглядно пояснить на графике. Произведение полусуммы скоростей в начале и в конце промежутка на величину промежутка времени есть площадь трапеции $ABCD$ (рис. 5): ее основания — AB и DC, а высота — AD; площадь трапеции равна

$$\frac{AB+DC}{2}\,AD=\frac{v(t_1)+v(t_2)}{2}(t_2-t_1).$$

Поэтому определение пути по полусуммам скоростей называется *методом трапеций*. При том виде, который имеет кривая $v(t)$ на рис. 5, площадь трапеции несколько больше площади, ограниченной прямыми BA, AD, DC и дугой BC кривой $v=v(t)$. Разность площади трапеций и площади, ограниченной дугой кривой, равна площади луночки (заштрихована на рис. 5), ограниченной хордой BC и дугой BC кривой. Сумма площадей всех таких луночек и дает ошибку — разность между истинным значением пути и вычисленным по методу трапеций. Сравнение с рис. 2 и 3 наглядно показывает, что при использовании метода трапеций ошибка, вообще говоря, будет меньше, чем при расчете по формулам типа (4), (5) (по «методу ступенек», или, как чаще говорят математики, *по методу прямоугольников*).

Разумеется, при сопоставлении пути и площади, связанной с графиком зависимости $v=v(t)$, необходимо учитывать масштаб, в котором составлен график (ср. со сказанным по этому поводу в § 1.7). Пусть на графике отрезок в 1 см по оси абсцисс соответствует промежутку времени в T с, а отрезок в 1 см по оси ординат соответствует скорости V см/с. Тогда при движении

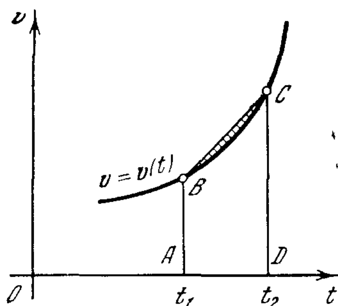


Рис. 3.1.5

с постоянной скоростью v_0 в течение времени от $t=a$ до $t=b$ путь будет равен $v_0(b-a)$, а площадь заштрихованного на рис. 1 прямоугольника равна (в см²)

$$S=\frac{v_0}{V}\frac{b-a}{T}.$$

Таким образом, здесь $z(a, b)=SVT$.

Это соотношение между пройденным путем и отвечающей ему площадью, ограниченной кривой $v=v(t)$, осью абсцисс и вертикалями $t=a$ и $t=b$, очевидно, сохраняет силу и в случае переменной скорости v, т. е. произвольной зависимости $v=v(t)$.

§ 2. Определенный интеграл

В предыдущем параграфе мы рассмотрели две задачи — задачу определения пройденного пути и, как оказалось, равносильную ей задачу определения площади под кривой; они привели к рассмотрению сумм особого вида с большим числом малых слагаемых. Строгая постановка этих задач приводит к понятию интеграла.

Величина $z=z(a, b)$ пути, найденная по заданной скорости $v(t)$, называется **определенным интегралом** функции $v(t)$ (скорости) по переменной t (времени), взятым в пределах от a до b (иногда говорят также «интеграл от функции $v(t)$»).

Дадим математическое определение интеграла, соответствующее тем идеям, которые были проиллюстрированы численным примером предыдущего параграфа. Это определение останется верным и в том случае, если рассматриваются не скорость и путь, а какие угодно другие физические или математические величины.

Итак, пусть дана функция $v=v(t)$. Для нахождения ее интеграла в промежутке от a до b мы разбиваем этот промежуток на большое число n малых промежутков. Значения аргумента t на границах малых промежутков обозначим $t_0, t_1, t_2, \ldots, t_{n-1}, t_n$, где, очевидно, $t_0=a$ и $t_n=b$ (рис. 1). Длины Δt малых промежутков времени[1] равны разностям соседних значений t. Таким образом, для какого-нибудь произвольного l (где $l=1, 2, \ldots, n$)

$$\Delta t_l=t_l-t_{l-1}.$$

Значки снизу у величин t и Δt представляют собой номера, или «индексы», значений t и промежутков Δt (см. сноску 1 в § 1.2).

Приближенное значение интеграла $z(a, b)$ дается формулой

$$z(a, b)\approx\sum_{l=1}^{l=n} v(t_{l-1})\Delta t_l. \qquad (3.2.1)$$

Запись $\sum\limits_{l=1}^{l=n}$ означает, что стоящее справа от этого знака[2] и зависящее от

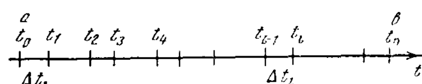


Рис. 3.2.1

индекса («номера») l выражение надо выписать при всех значениях l от 1 до n и все эти выражения надо сложить. Так, например, если $n=10$, то

$$\sum_{l=1}^{l=10} v(t_{l-1})\Delta t_l=$$
$$=v(t_0)\Delta t_1+v(t_1)\Delta t_2+\ldots+v(t_9)\Delta t_{10}.$$

В примере из § 1 (см. таблицу на с. 79) $t_0=1$; $t_1=1{,}1$; $t_2=1{,}2$; ... и

$$z(1, 2)\approx\sum_{l=1}^{l=10} t_{l-1}^2\Delta t_l=2{,}185.$$

В приближенном выражении (1) в каждом промежутке значение функции $v(t)$ бралось в начале промежутка — в точке t_{l-1}. Другое приближенное выражение получим, беря в каждом промежутке значение функции в конце промежутка:

$$z(a, b)\approx\sum_{l=1}^{l=n} v(t_l)\Delta t_l. \qquad (3.2.2)$$

В примере из § 1 при $n=10$ эта сумма равнялась 2,485.

---

[1] Если разбить промежуток от a до b на n равных частей, то каждый промежуток $\Delta t=\frac{b-a}{n}$. Однако вовсе не обязательно, чтобы все части, на которые разбит промежуток, были одинаковы — нужно лишь, чтобы каждый промежуток Δt был мал. Читатель может убедиться в этом, продумав пример путь—скорость из § 1; см. также упр. 3.

[2] Знак Σ есть прописная греческая буква сигма, символизирующая сложение.

Определенным интегралом функции $v(t)$, взятым в интервале от a до b, называется п р е д е л, к которому стремятся суммы (1) и (2) при стремлении всех промежутков Δt_l к нулю. Обозначается интеграл так:

$$z(a, b) = \int_a^b v(t)\,dt \tag{3.2.3}$$

(читается: «зет от а, бэ равно интегралу от а до бэ, вэ от тэ де тэ»). Знак $\int$ («интеграл») происходит от латинской буквы S (первой буквы слова «сумма»): он получился растягиванием буквы S в вертикальном направлении.

Запись dt (вместо Δt) в правой части (3) означает, что для получения точного значения интеграла необходимо перейти к пределу, когда все промежутки Δt стремятся к нулю (подобно тому как производная $\frac{dz}{dt}$ получается из отношения $\frac{\Delta z}{\Delta t}$, если устремить Δz и Δt к нулю и перейти к пределу). Формулы (1) и (2), в которых величины Δz малы, но конечны, дают только п р и б л и ж е н н о е значение интеграла, подобно тому как отношение $\frac{\Delta z}{\Delta t}$ с конечными Δt и Δz дает лишь приближенное значение производной.

Когда малые промежутки Δt становятся все мельче и мельче, то уже становится безразличным, брать ли значение функции v в начале, в конце или где-нибудь внутри промежутка, т. е. безразлично, исходить ли из (1), из (2) или из какого-либо иного выражения «интегральной суммы», подобной тем, которые фигурируют в правых частях (1) и (2) (можно, например, суммировать слагаемые $v(t_{\text{ср}})\,\Delta t$, где $t_{\text{ср}}$ — середина соответствующего промежутка Δt). Поэтому в записи (3) стоит просто $v(t)$, т. е. значение функции v в промежутке Δt без конкретизации того, как именно выбирается значение t внутри (или на концах) рассматриваемого промежутка.

Еще одно отличие интеграла (3) от приближающих его сумм (1) и (2) заключается в том, что при уменьшении величин Δt и при увеличении числа малых промежутков мы отказываемся от того, чтобы нумеровать их. Поэтому у интеграла указываются только пределы изменения t от a до b. Величина a ставится внизу и называется *нижним пределом* интегрирования, величина b стоит у верхнего конца знака интеграла и называется *верхним пределом* интегрирования [3]. Промежуток изменения t от a до b называется *промежутком интегрирования*. Функцию $v(t)$ в выражении интеграла называют *подынтегральной функцией*, t — *переменной интегрирования*.

Таким образом, интеграл определяется как предел, к которому стремится сумма произведений значений функции на разность соответствующих значений аргумента при стремлении к нулю всех разностей аргумента [4]:

$$\lim_{\Delta t_l \to 0} \sum_{l=1}^{l=n} v(t_l)\,\Delta t_l = \lim_{\Delta t_l \to 0} \sum_{l=1}^{l=n} v(t_{l-1})\,\Delta t_l = \int_a^b v(t)\,dt. \tag{3.2.4}$$

Хотя первая и вторая суммы в равенстве (4) при конечном числе l малых промежутков Δt_l различны, пределы их при неограниченном уменьшении всех промежутков Δt совпадут, — и эти-то пределы и дают численное значение интеграла.

При стремлении Δt к нулю каждое отдельное слагаемое стремится к нулю, но зато возрастает, стремится к бесконечности число членов суммы. Сама сумма стремится к вполне определенному пределу, являющемуся решением рассматриваемой нами задачи и называемому *интеграл*. Если функция представляет собой мгновенную скорость, то этот предел, т. е. интеграл от рассматриваемой функции, равен пройденному пути. Если подынтегральная функ-

---

[3] В этом параграфе слово «предел» употребляется в двух разных смыслах. Во-первых, интеграл есть предел суммы в том же смысле, в котором производная есть предел отношения: здесь слово «предел» соответствует знаку lim. Кроме того, мы говорим о пределах изменения t от a до b, т. е. о *пределах интегрирования* a и b. Здесь слово «предел» имеет уже совсем другой смысл. Внимательный читатель не запутается в этих двух разных значениях одного и того же слова.

[4] Пожалуй, точнее было бы писать $\lim\limits_{\Delta t_1, \Delta t_2, \ldots, \Delta t_n \to 0} \sum\limits_{l=1}^{l=n} v(t_l)\,\Delta t_l$ (и аналогично для второй суммы), но такая запись слишком уж громоздка!

ция указывает ординаты точек графика $y=f(x)$, то интеграл $\int\limits_a^b f(x)\,dx$ равен площади фигуры, ограниченной нашим графиком, осью x и вертикалями $x=a$ и $x=b$.

Разумеется, далеко не всякая сумма большого числа n малых слагаемых стремится к определенному пределу при $n\to\infty$: так, скажем, сумма n слагаемых, каждое из которых равно $\frac{1}{\sqrt{n}}$, равна, очевидно, $\sqrt{n}$ и при $n\to\infty$ неограниченно возрастает, а ее предел не может иметь никакого конечного значения. Почему же в нашем случае такой предел обязательно должен существовать?

Разобьем отрезок длины $b-a$ на n равных промежутков длины $\frac{b-a}{n}$. Если для простоты взять скорость v постоянной, то мы получим сумму n слагаемых, каждое из которых равно $v\Delta t=v\frac{b-a}{n}$. В итоге вся сумма (т. е. пройденный путь) будет равна $nv\Delta t=\frac{nv\,(b-a)}{n}=v\,(b-a)$, т. е. она не зависит от n. Здесь очень важно, что каждое отдельное слагаемое убывает как раз в таком отношении $\left(\text{пропорционально } \frac{1}{n}\right)$, в каком растет и число n слагаемых. Нетрудно понять, что в случае переменной скорости и разбиения всего интервала длины $b-a$ на n равных малых отрезков имеет место примерно то же обстоятельство — убывание самих слагаемых обратно пропорционально росту их числа. Так, при разбиении каждого маленького отрезка длины Δt на две равные части $\Delta_1 t$ и $\Delta_2 t$, отвечающие этим половинкам слагаемые интегральной суммы окажутся примерно вдвое меньше первоначального, «большого» слагаемого; но и общее число слагаемых тут удваивается, так что порядок величины всей суммы не изменится [5]. Читателю, которого это рассуждение еще не убедило, мы рекомендуем проделать упражнения к этому параграфу.

Это разъяснение полезно, если исходить из математического определения интеграла как предела суммы. В физических же задачах существование интеграла, т. е. предела рассматриваемых «интегральных сумм», как правило, очевидно. Например, очевидно, что тело, движущееся с конечной (постоянной или непостоянной) скоростью за конечное время пройдет вполне определенный (конечный) путь. Дальше мы подробно расскажем, как с помощью интеграла вычислить площадь криволинейной фигуры (выше мы бегло уже касались этого вопроса). Здесь также, очевидно, не вызывает никакого сомнения само существование ответа задачи, т. е. площади, а значит, и существование интеграла.

Поскольку переменная интегрирования может принимать значения a и b, то ясно, что пределы интегрирования являются размерными величинами: их размерность совпадает с размерностью переменной интегрирования (в примере путь—скорость пределы интегрирования имеют размерность в р е м е н и). (Относительно размерности всего интеграла см. конец этого параграфа.)

Ясно, что величина определенного интеграла зависит лишь от значений подынтегральной функции в н у т р и промежутка интегрирования; значения функции вне этого промежутка на величину интеграла влиять никак не могут и нас нисколько не интересуют. Так, пройденный путь $z\,(a,\ b)$ зависит, конечно, лишь от значения скорости $v=v\,(t)$ внутри промежутка интегрирования, но совсем не зависит от того, какой была скорость раньше, до момента $t=a$, и какой она будет потом (после момента $t=b$).

В § 1 отмечалось, что путь можно определить, вычислив площадь на графике зависимости скорости от времени. Задача нахождения п л о щ а д и S, ограниченной сверху линией с заданным уравнением $y=y\,(x)$, снизу — осью абсцисс (осью x), с боков — линиями $x=a$ и $x=b$, т. е. площади *криволиней-*

[5] Заметьте, что в примере с n слагаемыми по $\frac{1}{\sqrt{n}}$ удвоение общего числа слагаемых сопровождается уменьшением каждого из слагаемых всего лишь в $\frac{1}{\sqrt{2}}$ раза. В примере со слагаемыми, равными $\frac{1}{n^2}$, та же операция приводит к четырехкратному уменьшению каждого слагаемого, так что новая сумма составляет лишь половину первоначальной.

ной трапеции $ABDC$, основаниями которой служат параллельные отрезки AC и BD прямых $x=a$ и $x=b$, а боковыми сторонами — отрезок AB оси абсцисс и дуга CD линии $y=y(x)$ (рис. 2), сводится к вычислению интеграла

$$S=\int_a^b y(x)\,dx. \qquad (3.2.5)$$

Чтобы убедиться в этом, достаточно снова обратиться к рис. 1.2, где мы теперь считаем, что по оси абсцисс откладываются значения (произвольной!) независимой переменной x, а по оси ординат — значения функции $y=y(x)$ этой переменной, причем величины x и y рассматриваются как расстояния переменной точки $M=M(x, y)$ от осей координат (см. § 1.2) вне всякой связи с такими физическими понятиями, как время или путь. Сумма площадей заштрихованных на рис. 1.2, *а* прямоугольников будет в этом случае равна $\sum\limits_{l=1}^{n} y(x_{l-1})\,\Delta x_l$, а такая же сумма, изображенная на рис. 1.2,*б*, — $\sum\limits_{l=1}^{l=n} y(x_l)\,\Delta x_l$.

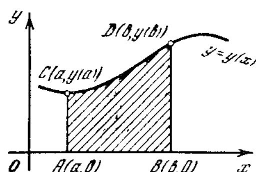


Рис. 3.2.2

В пределе при $\Delta x_l \to 0$ эти суммы, по определению, обращаются в интеграл, а сумма площадей прямоугольников стремится к площади, ограниченной кривой $y(x)$, так как чем меньше все Δx_l, тем ближе к кривой ломаная (зубчатая) линия, ограничивающая прямоугольники (ср. рис. 1.3).

В заключение отметим, что определенный интеграл зависит от подынтегральной функции и пределов интегрирования, но не зависит от обозначения переменной интегрирования. В самом деле, пусть дана, скажем, подынтегральная функция

$$v(t)=3t^2+5.$$

Заменив букву t буквой x, получим

$$v(x)=3x^2+5.$$

При вычислении интеграла безразлично, как называлась переменная интегрирования, важно только, в каких пределах она менялась, какова была функция этой переменной. Поэтому

$$z(a, b)=\int_a^b v(t)\,dt=\int_a^b v(x)\,dx=$$

$$=\int_a^b v(u)\,du=\int_a^b v(\lambda)\,d\lambda$$

или даже (конечно, такая запись никогда не применяется!)

$$z(a, b)=\int_a^b v(щ)\,dщ.$$

Переменную интегрирования можно обозначить (назвать) как угодно — это не влияет на результат!

Переменная, которая фигурирует лишь в промежуточных выкладках, а в окончательный результат не входит, называется *немой* переменной. Таким образом, переменная интегрирования является немой. Обычную, не немую, переменную можно заменять другой лишь одновременно во всех частях формулы; например, соотношение $(x+1)^2=x^2+2x+1$ нельзя переписать так: $(x+1)^2=t^2+2t+1$. В интегралах же можно писать $z(a, b)=\int_a^b v(t)\,dt$ или $z(a, b)=\int_a^b v(x)\,dx$.

Наконец, обратимся к вопросу о размерности интеграла. По определению,

$$\int_a^b y(x)\,dx=\lim_{\Delta x_l\to 0}\sum_{l=1}^{l=n} y(x_l)\,\Delta x_l. \qquad (3.2.6)$$

При этом, если величина x измеряется в единицах e_1, а величина y — в единицах e_2, то каждое слагаемое стоящей в правой части (6) суммы имеет размерность e_1e_2 (множитель Δx_l имеет размерность e_1, а множитель $y(x_l)$ — размерность e_2), поэтому ту же размерность имеет и предел суммы, т. е. интеграл. Так, если скорость $v(t)$ измеряется в см/с, а время t — в с, то путь $z(a, b)=\int_a^b v(t)\,dt$ измеряется в единицах (см/с)·с=см; если абсцисса x и

ордината y измеряются в сантиметрах, то площадь $S=\int_a^b y(x)\,dx$ измеряется в квадратных сантиметрах (см·см=см²), и т. д. При этом переход к новым единицам e_1' измерения x и к новым единицам e_2' измерения y приводит к умножению интеграла (5) на множитель k_1k_2, где $e_1=k_1e_1'$ и $e_2=k_2e_2'$; при этом новый (в k_1k_2 раз больший) интеграл выражает ту же величину, что и раньше, только измеренную в новых единицах $e_1'e_2'$. Так, переход от сантиметров к миллиметрам увеличивает численное значение площади $S=\int_a^b y(x)\,dx$ в $10\cdot 10=100$ раз, а переход от сантиметров в секунду к метрам в минуту (и от секунд к минутам) приводит к необходимости умножить (ранее выраженное в сантиметрах) значение пути $z(a, b)=\int_a^b v(t)\,dt$ на $\left(0{,}01\Big/\frac{1}{60}\right)\cdot\frac{1}{60}=0{,}01$ (новое значение пути, разумеется, будет уже выражено в метрах).

Упражнения

3.2.1. Рассмотрите случай $v=kt+s$ (равноускоренное движение). Найдите путь за время от a до b, разбивая этот промежуток времени на m равных интервалов; воспользуйтесь тем, что слагаемые сумм образуют арифметическую прогрессию. Найдите предел суммы при $m\to\infty$. Сравните полученное выражение с площадью трапеции на плоскости t, v, равной пройденному пути.

3.2.2. Рассмотрите случай $v=t^2$ и для этого случая найдите путь за время от $t=1$ до $t=2$, т. е. найдите интеграл $\int_1^2 t^2dt$. Для этого можно воспользоваться разбиением интервала от 1 до 2 на m равных частей и сосчитать сумму $\sum t_{i-1}^2\frac{1}{m}$ или $\sum t_i^2\frac{1}{m}$. Сравните эти две суммы. [Указание. Для вычисления указанных сумм воспользуйтесь формулой

$$1^2+2^2+3^2+\ldots+n^2=\frac{n(n+1)(2n+1)}{6}.\Big]$$

3.2.3. Вычислите $\int_1^2 x^3dx$, разбив отрезок от 1 до 2 на m меньших отрезков Δx, длины которых составляют геометрическую прогрессию, и перейдите затем к пределу при $m\to\infty$.

§ 3. Связь между интегралом и производной

Производная от интеграла. В предыдущих параграфах мы рассмотрели отдельно понятия производной и интеграла. В настоящем параграфе мы (на примере путь—скорость) установим связь между этими двумя понятиями, которая обращает исчисление производных и исчисление интегралов в одну содержательную и богатую приложениями науку: *дифференциальное и интегральное исчисление*, или *математический анализ*.

Будем считать заданной (известной) зависимость $v=v(t)$ мгновенной скорости v от времени t. Момент $t=a$ начала пути мы зафиксируем раз и навсегда (поезд отправляется в путь в момент $t=a$ с известного нам вокзала); при этом проходимый за время от a до b путь будет зависеть от момента времени b, будет являться функцией b. Для того чтобы подчеркнуть это обстоятельство, мы условимся обозначать пока «конечный» (неопределенный!) момент времени не буквой b, а, скажем, буквой u — ведь переменные величины мы привыкли (хоть это, разумеется, лишь условность!) обозначать последними буквами латинского алфавита:

$$z(a, u)=\int_a^u v(t)\,dt. \tag{3.3.1}$$

Мы знаем, что

$$z(a, u)=z(u)-z(a) \tag{3.3.2}$$

(в примере с поездом постоянную $z(a)$ уместно даже считать нулем). Продифференцируем левую и правую части (2) (т. е. возьмем производные от обеих частей); здесь мы считаем величину u (независимой) переменной, а a — постоянной, поэтому слагаемое $-z(a)$ в правой части — постоянное и $\frac{d(-z(a))}{du}=0$. Мы получим

$$\frac{dz(a, u)}{du}=\frac{dz(u)}{du}.$$

Но ведь производная от координаты тела по времени есть не что иное, как

мгновенная скорость тела; поэтому $\frac{dz(u)}{du}=v(u)$ — скорость в момент u и, значит,

$$\frac{dz(a,\ u)}{du}=v(u). \qquad (3.3.3)$$

Подставим сюда выражение (1) пути $z(a,\ u)$ в виде интеграла; получим

$$\frac{d}{du}\left(\int_a^u v(t)\,dt\right)=v(u). \qquad (3.3.4)$$

Это равенство является важнейшим общим свойством определенного интеграла. В таком виде оно является общей математической теоремой; его правильность не зависит от того, является ли $v(t)$ скоростью (а интеграл — путем) или $v(t)$ есть какая-то другая величина. Для любой функции, скажем $y(x)$, имеем:

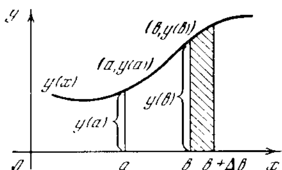


Рис. 3.3.1

$$\frac{d}{db}\left(\int_a^b y(x)\,dx\right)=y(b), \qquad (3.3.4а)$$

где теперь мы снова обозначаем верхний (переменный) предел интегрирования через b.

Словесная формулировка теоремы такова: *производная от определенного интеграла по его верхнему пределу равна значению подынтегральной функции от верхнего предела интеграла.*

Ввиду важности этой теоремы (ее называют теоремой Ньютона—Лейбница[1]) дадим другой вывод ее, основанный на рассмотрении площади. Производную будем вычислять по всем правилам, как предел отношения приращения функции к приращению независимой переменной. Итак, пусть

$$I(a,\ b)=\int_a^b y(x)\,dx.$$

Этот интеграл есть площадь, ограниченная сверху кривой $y(x)$, снизу — осью x, а слева и справа — вертикалями $x=a$ и $x=b$ (см. рис. 2.2).

Как найти приращение ΔI интеграла I при малом изменении Δb верхнего предела b? По определению приращения,

$$\Delta I=I(a,\ b+\Delta b)-I(a,\ b).$$

Но площадь, равная интегралу $I(a,\ b+\Delta b)$, отличается от площади $I(a,\ b)$ тем, что правая вертикаль сместилась вправо еще на (малое) Δb (ср. рис. 1 и рис. 2.2). Следовательно, приращение ΔI есть разность двух площадей: площади от a до $b+\Delta b$ и площади от a до b. Ясно, что ΔI есть площадь тонкой полоски, заштрихованной на рис. 1, основанием которой на оси x (или проекцией которой на ось x) является отрезок длины Δb.

Искомая производная равна пределу

$$\frac{dI(a,\ b)}{db}=\lim_{\Delta b\to 0}\frac{\Delta I}{\Delta b}.$$

Но, очевидно, что при стремлении Δb к нулю площадь полоски ΔI приближается к произведению $y(b)\,\Delta b$ — ее ширины Δb на высоту, а отношение $\frac{\Delta I}{\Delta b}$ — к величине $y(b)$.

Таким образом, мы снова наглядно убедились в справедливости теоремы Ньютона—Лейбница:

$$\frac{d}{db}\left(\int_a^b y(x)\,dx\right)=y(b).$$

Определенный интеграл известной функции $y(x)$ или $v(t)$ есть функция пределов a, b интегрирования — функция двух переменных a и b. Формула (4) (или (4а)) указывает значение производной этой функции по верхнему пределу интегрирования — по переменной b; ниже, в § 5, мы найдем также значение ее производной по переменной a[2].

Данное в предыдущем параграфе определение интеграла как предела суммы объясняет роль этого важней-

---

[1] В некоторых старых учебниках дифференциального и интегрального исчисления теорему Ньютона—Лейбница называли *основной теоремой высшей математики*; это название звучит несколько выспренно, но оно вполне оправдано.

[2] По существу здесь, конечно, речь идет о частных производных функции $I(a,b)=\int_a^b y(x)\,dx$ двух переменных a и b (см. § 4.12).

шего понятия в решении физических или геометрических задач: в вычислении пройденного пути по известной (переменной) скорости, в определении площади криволинейной трапеции и в других проблемах, приводящих к составлению «интегральных сумм» (в дальнейшем нам встретится много примеров таких проблем). Но это определение не дает удобного общего способа вычисления интеграла или способа нахождения интеграла в виде формулы как функции пределов интегрирования [3].

Способ нахождения такой формулы следует из доказанной выше теоремы о производной от интеграла. При этом, кроме свойства производной от интеграла, используется еще второе свойство определенного интеграла: *определенный интеграл равен нулю, когда верхний и нижний пределы его совпадают*:

$$z(a, a) = \int_a^a v(t)\,dt = 0. \tag{3.3.5}$$

Это свойство вполне очевидно — путь равен нулю, если равно нулю время пути.

Сама формула, дающая значение интеграла как функцию пределов интегрирования, будет выведена таким способом в § 4.

Интеграл от производной. В выражении

$$\int_a^b v(t)\,dt \tag{3.3.6}$$

подынтегральная функция $v(t)$ может быть какой угодно, и регулярного способа, гарантирующего нахождение интеграла при любой функции $v(t)$, просто не существует. Пусть, однако, нам повезло — мы сумели найти такую функцию F, что подынтегральная функция $v(t)$ является производной этой функции F:

$$\frac{dF(t)}{dt} = v(t). \tag{3.3.7}$$

В этом наиболее благоприятном случае дело обстоит совсем хорошо: здесь легко найти точное значение интеграла. Для того чтобы выписать его, вспомним приближенное выражение приращения функции F (ср. формулу (2.4.6)):

$$\Delta F \approx F'(t)\,\Delta t = v(t)\,\Delta t,$$

или, в более обычных обозначениях, относящихся к интегралу $\int_a^b f(x)\,dx$, где $f(x) = \frac{dF(x)}{dx}$,

$$\Delta F \approx F'(x)\,\Delta x = f(x)\,\Delta x. \tag{3.3.8}$$

Величина, стоящая в правой части равенства (8), представляет собой как

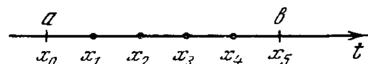


Рис. 3.3.2

раз одно из тех слагаемых, сумма которых определяет интеграл. Поэтому, положив, скажем, $\Delta x_l = x_l - x_{l-1}$ и, значит, $\Delta F = F(x_l) - F(x_{l-1})$, можно приближенно написать

$$\Delta F = F(x_l) - F(x_{l-1}) \approx f(x_l)(x_l - x_{l-1}) = f(x_l)\,\Delta x_l. \tag{3.3.9}$$

Равенство (9) является приближенным; однако чем меньше разность между x_l и x_{l-1}, т. е. приращение Δx_l, тем оно точнее. Но при уменьшении разности $x_l - x_{l-1} = \Delta x_l$, т. е. при сближении x_l и x_{l-1}, уменьшается и различие между $f(x_l)$ и $f(x_{l-1})$; поэтому в правой части формулы (9) с одинаковым правом, с одинаковой степенью точности можно написать и $f(x_{l-1})$, и $f(x_l)$ — мы выбрали последний вариант.

Выпишем теперь подобные (9) формулы для в с е х промежутков, на которые разбита область интегрирования, т. е. интервал от a до b. Пусть, например, интервал разбит на пять промежутков (рис. 2), так что $x_0 = a$, $x_5 = b$; тогда

$$\begin{aligned} F(x_1) - F(x_0) &\approx f(x_1)(x_1 - x_0), \\ F(x_2) - F(x_1) &\approx f(x_2)(x_2 - x_1), \\ F(x_3) - F(x_2) &\approx f(x_3)(x_3 - x_2), \\ F(x_4) - F(x_3) &\approx f(x_4)(x_4 - x_3), \\ F(x_5) - F(x_4) &\approx f(x_5)(x_5 - x_4). \end{aligned} \tag{3.3.10}$$

[3] Лишь в редких случаях и с трудом удается провести суммирование произвольного числа малых слагаемых, фигурирующих в определении интеграла (см., например, упр. 2.2 и 2.3).

Сложим эти пять равенств. В левой части сократятся все значения функции F при промежуточных значениях x, останется только

$$F(x_5)-F(x_0)=F(b)-F(a).$$

В правой части получаются как раз такие суммы, с помощью которых мы в § 1 приближенно выражали интеграл (там мы выражали путь $z(a,\ b)$ при данной скорости $v(t)$). Итак,

$$F(b)-F(a)\approx\sum_{l=1}^{l=5}f(x_l)(x_l-x_{l-1})=$$

$$=\sum f(x_l)\,\Delta x_l\approx\int_a^b f(x)\,dx,$$

где $f(x)=\frac{dF}{dx}$.

Чем меньше каждое приращение Δx, т. е. величина x_l-x_{l-1}, тем точнее выражение (9) приращения ΔF. Но при уменьшении всех разностей $\Delta x_l=x_l-$ $-x_{l-1}$ (и при неограниченном увеличении числа отрезков Δx_l) суммы стремятся к интегралу. Поэтому равенство

$$F(b)-F(a)=\int_a^b f(x)\,dx,\quad\text{где}\quad f(x)=\frac{dF}{dx},\tag{3.3.11}$$

является уже совершенно точным.

Последнее утверждение представляется достаточно убедительным, однако его можно обосновать и более точным подсчетом. Дело в том, что ошибка, допускаемая каждой подобной (8) или (9) формулой, вообще говоря, является *квадратичной*: она примерно пропорциональна второй степени длины Δt интервала, т. е. имеет вид $c(\Delta t)^2$, где число $c=c(t)$ от t зависит достаточно слабо — с вполне удовлетворительной степенью точности его можно считать постоянным (все это будет подробно объяснено в гл. 6). Таким образом, при $\Delta x=\frac{b-a}{n}$ в каждом из выражений (10) (число которых теперь надо считать равным не 5, а n) допускается ошибка, пропорциональная $\Delta t^2=\frac{1}{n^2}(b-a)^2$. Поэтому полная ошибка, накапливаемая при суммировании таких приближенных равенств, будет пропорциональна $n\left[\frac{1}{n^2}(b-a)^2\right]=\frac{1}{n}(b-a)^2$, откуда и вытекает, что она неограниченно уменьшается при $n\to\infty$.

Формула (11) устанавливает связь между задачами об интеграле и о производной. Из этой формулы следует, что если бы удалось найти такую функцию F, производная которой равна подынтегральной функции f, то задача вычисления интеграла была бы решена — нам осталось бы только принять за аргумент этой функции значения a и b и найти разность

$$F(b)-F(a).$$

Ввиду огромной важности формулы (11) в следующих параграфах мы дадим другой ее вывод на основе более подробного рассмотрения свойств интеграла и функции F.

§ 4. Неопределенный интеграл

В предыдущих параграфах мы ввели понятие определенного интеграла как предела суммы большого числа малых слагаемых. В § 3 было выявлено главное свойство определенного интеграла: производная определенного интеграла по верхнему пределу интегрирования равна подынтегральной функции:

$$\text{если}\quad z(a,\ b)=\int_a^b v(t)\,dt,\quad\text{то}\quad\frac{dz(a,\ b)}{db}=v(b).\tag{3.4.1}$$

Теперь мы хотим воспользоваться этим свойством для вычисления определенного интеграла.

Итак, будем искать такую функцию переменной b, производная которой есть известная функция $v(b)$. Обозначим эту функцию через $F(b)$. По определению,

$$\frac{dF(b)}{db}=v(b).\tag{3.4.2}$$

Равенство (2) не определяет полностью функцию $F(b)$. Ведь прибавление любой константы к функции не меняет ее производную; следовательно, если $F(b)$ удовлетворяет (2), то и функция $G(b)=F(b)+C$ при любой постоянной C удовлетворяет тому же равенству.

Функцию $F(b)$, удовлетворяющую условию (2), называют *неопределенным интегралом* функции $v(b)$; аналогично, если

$$\frac{dF(x)}{dx}=f(x),\tag{3.4.2a}$$

то говорят, что функция $F(x)$ есть **неопределенный интеграл** функции $f(x)$. В этом названии отражаются два свойства величины (функции) F. Во-первых, функция F имеет такую же производную по b, как и (определенный) интеграл $z(a, b)$, или $\int_a^b f(x)\,dx$ (ср. (1) и (2) или (2а)); поэтому F также называют «интегралом». С другой стороны, эта функция условием (2) (или (2а)) полностью не определена, ибо к ней всегда можно прибавить произвольную постоянную C; поэтому здесь представляется вполне уместным прилагательное «неопределенный».

Любое решение уравнения (2) (или (2а)) может отличаться от одного какого-нибудь его решения $F(b)$ (или $F(x)$) только на ту или иную п о с т о я н н у ю величину. В самом деле, если $G(b)$ есть другое решение (2), то для разности $F-G$ получим

$$\frac{d}{db}[F(b)-G(b)]=v(b)-v(b)=0.$$

Но производная одной лишь постоянной функции равна нулю при любых значениях аргумента — только полностью покоящееся тело все время имеет нулевую скорость.

Определенный интеграл $z(a, b)$ согласно (1) тоже является одним из решений уравнения (2). Значит, и $z(a, b)$ можно представить в виде

$$z(a, b)=F(b)+C, \tag{3.4.3}$$

где $F(b)$ — какое-то решение (2), C — (неопределенная) постоянная, которую нам требуется определить. Для того чтобы добиться этого, воспользуемся вторым свойством определенного интеграла — равенством его нулю, когда верхний и нижний пределы интегрирования совпадают:

$$z(a, a)=0 \tag{3.4.4}$$

(см. (3.5)). Подставляя в (3) $b=a$ и используя (4), получим

$$0=F(a)+C, \quad \text{т. е.} \quad C=-F(a).$$

Отсюда окончательно получаем

$$z(a, b)=F(b)-F(a), \tag{3.4.5}$$

или

$$\int_a^b f(x)\,dx=F(b)-F(a), \quad \text{где } F'(x)=f(x). \tag{3.4.5а}$$

Отметим, что «неопределенность» функции $F(b)$ ничуть не мешает вычислению с ее помощью определенного интеграла по формуле (5), или (5а). В самом деле, возьмем вместо $F(b)$ какое-нибудь другое решение уравнения (2), например $G(b)$, отличающееся от $F(b)$ на постоянную величину C: $G(b)=F(b)+C$.

Будем вычислять определенный интеграл по формуле (5), заменив только F на G:

$$z(a, b)=G(b)-G(a)=[F(b)+C]-$$
$$-[F(a)+C]=F(b)-F(a).$$

Мы получили тот же результат, что и выше!

В примере путь—время неопределенный интеграл удобно обозначать той же буквой z, которой мы обозначаем определенный интеграл. При данной подынтегральной функции $v(t)$ определенный интеграл зависит от верхнего и от нижнего пределов интегрирования, т. е. является функцией $z=z(a, b)$ двух переменных a и b. Неопределенный интеграл есть функция одной переменной — обозначим ее t. Итак, неопределенный интеграл есть функция, удовлетворяющая условию

$$z'(t)=\frac{dz(t)}{dt}=v(t). \tag{3.4.6}$$

С помощью этой функции определенный интеграл $z(a, b)$ функции $v(t)$ находится по формуле

$$z(a, b)=\int_a^b v(t)\,dt=z(b)-z(a). \tag{3.4.7}$$

Принята следующая краткая запись разности значений одной и той же функции при двух различных значениях переменной:

$$z(t)\big|_a^b=z(b)-z(a). \tag{3.4.8}$$

В этой записи слева стоит функция немой переменной t, правее которой ставится вертикальная черта и у черты сверху то значение переменной, при котором мы хотим взять функцию со

знаком плюс, а снизу — то значение, при котором функция берется со знаком минус.

Подставляя в фигурирующий в (7) интеграл функцию $v(t)$, выраженную через $z(t)$ согласно (6), а в правую часть (7) — выражение (8), получим тождество

$$\int_a^b z'(t)\,dt = z(t)\Big|_a^b, \tag{3.4.9}$$

или, в других обозначениях,

$$\int_a^b F'(x)\,dx = F(x)\Big|_a^b. \tag{3.4.9a}$$

Обратите внимание на похожее расположение a и b слева и справа, облегчающее запоминание формулы (9) (или (9а)).

Мы достаточно подробно обсудили само понятие неопределенного интеграла; пора перейти к примерам. Рассмотрим задачу о пути, пройденном за время от a до b при скорости движения $v(t)=t^2$. Искомый путь равен определенному интегралу

$$z(a,\ b)=\int_a^b t^2 dt.$$

В этой задаче неопределенный интеграл $z(t)$ получается из решения уравнения

$$\frac{dz(t)}{dt}=v(t)=t^2. \tag{3.4.10}$$

Но мы знаем, что $\frac{d(t^3)}{dt}=3t^2$, значит, $\frac{d(t^3/3)}{dt}=\frac{1}{3}(3t^2)=t^2$. Следовательно, уравнению (10) удовлетворяет функция

$$z(t)=\frac{t^3}{3}+C, \tag{3.4.11}$$

где C — произвольная постоянная, которую можно принять и равной нулю — ведь, как мы знаем, от этого разность значений функции z при двух значениях независимой переменной нисколько не зависит.

Подставим (10) и (11) в формулу (9):

$$\int_a^b t^2 dt=\frac{t^3}{3}\Big|_a^b=\frac{b^3}{3}-\frac{a^3}{3}.$$

В частном случае $a=1$, $b=2$ имеем

$$\int_1^2 t^2 dt=\frac{8}{3}-\frac{1}{3}=\frac{7}{3}\approx 2{,}333.$$

Таким образом, с помощью неопределенного интеграла мы в нескольких строчках получили точно результат, к которому мы мучительно приближались в § 1 численными расчетами.

Определенный интеграл есть предел сумм вида

$$v(t_0)(t_1-t_0)+v(t_1)(t_2-t_1)+\ldots$$

при стремлении к нулю каждого слагаемого и соответствующем увеличении числа слагаемых; для его приближенного вычисления нужно разбить область интегрирования на несколько интервалов, найти приближенное значение пути $v\Delta t$ в каждом интервале и сложить их. Чтобы получить хорошую точность, нужно сделать много арифметических операций. Но если известен неопределенный интеграл $z(t)$, т. е. известна функция, производная которой равна подынтегральной функции $v(t)$, то любой определенный интеграл $\int_a^b v(t)\,dt$ получается немедленно по формуле (9). Умение находить функции с заданной производной (неопределенные интегралы) «неожиданно» дает мощный способ вычисления сумм (определенных интегралов).

Неопределенный интеграл иногда называют *первообразной функцией*, воспринимая этот термин как обратный к понятию «производная»: речь идет о той функции, от которой берется (известная нам) производная. Термин «первообразная» широко используется учебниками, которые начинают интегральное исчисление именно с неопределенных интегралов, а определенные интегралы вводят позже. Мы этот термин использовать не будем.

Неопределенный интеграл всегда можно выразить через определенный

$$z(t)=C+\int_{a_0}^t v(x)\,dx. \tag{3.4.12}$$

Применяя правило о производной определенного интеграла по верхнему пределу, легко проверить, что заданная формулой (12) функция $z(t)$ удовлетворяет (6) при любых постоянных C и a_0.

Во всех задачах в ответ всегда входит разность значений $z(b)-z(a)$, ко-

торая не зависит от C и a_0. Поэтому (12) можно записать короче:

$$z(t)=\int^t v(x)\,dx.$$

Наконец, часто пишут еще короче:

$$z(t)=\int v(t)\,dt. \tag{3.4.13}$$

Этот способ записи весьма употребителен, и мы тоже будем им пользоваться; однако надо иметь в виду, что равенство (13), в сущности, неправильно [1]. Его можно сравнить с теми грамматически или формально логически неправильными выражениями, которые широко применяются в разговорной речи, всем (кроме детей и педантов-придир) понятны и принимаются без возражений, вроде выражений «съесть еще одну тарелочку» или «плюнуть на это дело». В математике достаточно часто применяются не совсем точные, но достаточно понятные записи и выражения; обычно они никаких затруднений ни у кого не вызывают.

В записи (13) нарушено основное правило, согласно которому (немая!) переменная интегрирования не имеет права использоваться в окончательном результате. Употребляя краткую запись (13), надо всегда помнить, что это лишь условное сокращение точного выражения (12), в котором $C=z(a_0)$.

Известные нам формулы для производных дают первую таблицу неопределенных интегралов:

$$\int dt=t,\quad \int t\,dt=\frac{t^2}{2},\quad \int t^2dt=\frac{t^3}{3},$$

$$\int\frac{dt}{t^2}=-\frac{1}{t}.$$

Кроме того, результаты упр. 2.4.2 и 2.4.3, позволяют написать:

$$\int\frac{dt}{\sqrt{t}}=2\sqrt{t},\quad \int\sqrt{t}\,dt=\frac{2}{3}\sqrt{t^3}.$$

Для того чтобы проверить любую из этих формул, достаточно найти производную от правой части. Если при этом получится функция, стоящая под интегралом, то формула верна.

Подробно способы нахождения неопределенных интегралов от различных функций рассматриваются в гл. 5. Благодаря связи между интегралом и производной удается найти интегралы большого числа функций.

Задача интегрирования является технически более сложной, чем задача нахождения производных. Эта сложность проявляется, в частности, в том, что при интегрировании рациональных (не содержащих корней) алгебраических выражений в ответе иногда появляются логарифмы и обратные тригонометрические функции; при интегрировании алгебраических выражений с корнями результат может выражаться при помощи новых, не элементарных, функций, которые нельзя выразить с помощью конечного числа действий над алгебраическими, степенными и тригонометрическими функциями (ср. § 5.5). Однако трудности выражения интегралов не должны заслонять принципиальную простоту и ясность понятия интеграла. Если нельзя (или трудно) сосчитать интеграл по формуле (5а), то его всегда можно подсчитать приближенно при помощи хотя и трудоемких, но в принципе весьма простых расчетов.

В последнее время, в связи с появлением карманных компьютеров вычисление сумм 10, 20 или 50 членов стало достаточно простым делом, требующим 10—30 мин в зависимости от сложности подынтегрального выражения. Поэтому во многих случаях такое прямое вычисление оказывается предпочтительным по сравнению с получением сложных формул. В дневнике замечательного физика Энрико Ферми приведен расчет довольно простого определенного интеграла. Несмотря на то что соответствующий ему неопределенный интеграл выражался через функции арксинус и натуральный логарифм, Ферми предпочел найти интеграл численно, не пользуясь формулой (5а).

Наиболее быстрые и точные способы численного интегрирования рассматриваются во многих книгах. Кратко об этом рассказано, например, в книге [15].

---

[1] Заметьте, что в левой части (13) стоит какая-то *функция* $z(t)$, а в правой — неопределенный интеграл, т. е. *семейство* $F(t)+C$ *функций*, отличающихся одна от другой на константу (ибо в записи $F(t)+C$ число C произвольно).

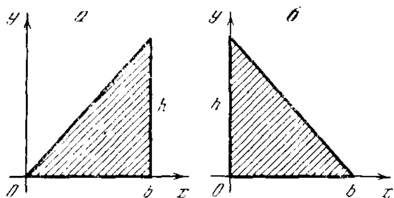


Рис. 3.4.1

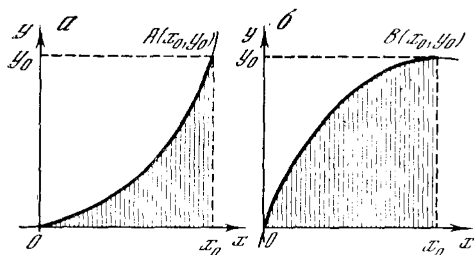


Рис. 3.4.2

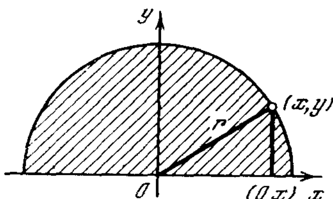


Рис. 3.4.3

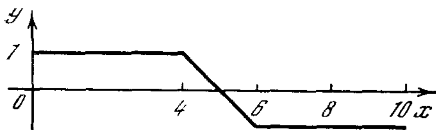


Рис. 3.4.4

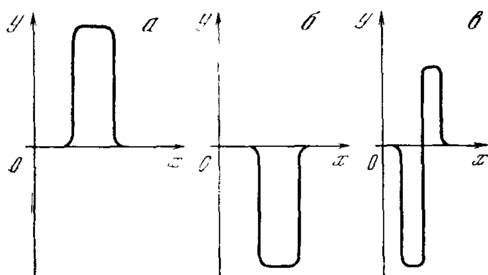


Рис. 3.4.5

Упражнения

3.4.1. Вычислите интегралы: а) $\int\limits_0^1 t^2dt$; б) $\int\limits_1^{1,1} t^2dt$; в) $\int\limits_1^2 \frac{dt}{t^2}$; г) $\int\limits_1^3 \frac{dt}{\sqrt{t}}$.

3.4.2. а) Найдите площадь прямоугольного треугольника с основанием b и высотой h при помощи интеграла. [Указание. Совместите начало координат с вершиной острого угла треугольника, а вершину прямого угла расположите на оси абсцисс (рис. 1, *а*). Найдите уравнение гипотенузы в этой системе координат и вычислите площадь как интеграл.]

б) Найдите площадь того же треугольника, совместив вершину прямого угла с началом координат, а вершину острого угла — с точкой $(b, 0)$ (рис. 1, *б*). [Указание. Воспользуйтесь очевидным свойством интеграла суммы двух функций: $\int (f+g)\,dx = \int f dx + \int g dx$.]

Замечание. Не возмущайтесь тем, что Вам приходится употреблять усилия для получения общеизвестного результата $S_\Delta = \frac{1}{2} bh$, — ведь методы интегрирования применимы и там, где элементарные способы нахождения площадей оказываются бессильными!

3.4.3. а) Найдите площадь под параболой $y=ax^2$ (рис. 2, *а*), ограниченную вертикалью $x=x_0$ и осью абсцисс; выразите площадь через координаты x_0, y_0 конца дуги OA параболы.

б) Тот же вопрос для случая параболы, проходящей через начало координат, горизонтальная касательная к которой проходит через ее точку $B(x_0, y_0)$ (рис. 2, *б*). [Указание. Ответ можно получить немедленно, пользуясь результатом упр. 3а. Тем не менее не поленитесь и сделайте все по порядку, не стремясь сократить работу с помощью остроумной находки. Уравнение параболы следует искать в виде $y=kx^2+px+q$, где величины k, p, q можно найти из условий прохождения линии через точки (x_0, y_0) и начало координат $(0, 0)$ и из условия горизонтальности касательной к параболе в точке (x_0, y_0). Площадь выразите через x_0, y_0. Если Вы не можете использовать результат упр. 3а, то сперва поступите как указано выше — и полученный результат сам подскажет связь между упр. а и б.]

3.4.4. Запишите выражение площади полукруга радиуса r (рис. 3) в виде определенного

интеграла. [Указание. Воспользуйтесь уравнением окружности: $x^2+y^2=r^2$.]

3.4.5. Найдите величину интеграла

$\int\limits_0^1 \frac{dx}{1+x^2}$ по формуле трапеций (см. § 1),

приняв значения $n=5$ и $n=10$. Вычисления ведите с четырьмя знаками после запятой.

Замечание. Точное значение этого интеграла равно $\pi/4=0{,}785398$ (см. далее упр. 5.2.4); приближенное вычисление интеграла дает возможность найти приближенное значение числа π.

3.4.6. Постройте график функции $F(x)=\int\limits_a^x y(x)\,dx$, где график функции $y(x)$ изображен на рис. 4. Значения a примите равными 0; 4; 8.

3.4.7. Постройте график функции

$$F(x)=\int\limits_0^x y(x)\,dx,$$

где графики (возможные) функции $y(x)$ изображены на рис. 5.

3.4.8. Постройте кривые

$$F(x)=\int\limits_0^x \varphi(x)\,dx,$$

графики функции $\varphi(x)$ (возможные) изображены на рис. О.6 и О.7 (см. с. 484). Сравните $F(x)$ с функциями $y(x)$, графики которых изображены на рис. 2.5.5 и 2.5.6.

§ 5. Свойства интегралов

Мы рассматривали выше самый простой случай определенного интеграла, когда подынтегральная функция положительна и верхний предел интегрирования больше нижнего:

$$z(a,\ b)=\int\limits_a^b v(t)\,dt,\quad v(t)>0 \text{ и } b>a.$$

В этом случае интеграл, очевидно, положителен, так как он равен пределу суммы положительных слагаемых. Интеграл имеет простой физический смысл пути (если $v=v(t)$ — это скорость) или площади (если $v=v(t)$ — уравнение кривой). А каким будет знак интеграла от отрицательной функции, т. е. в случае $v(t)<0$?

Оставим пока в силе условие $b>a$. В выражении суммы (переходящей в пределе в интеграл) в каждом члене множитель Δt положителен, множитель $v(t)$ отрицателен, каждое слагаемое отрицательно, сумма отрицательна, интеграл также отрицателен. Итак, если $v(t)<0$ при $a<t<b$ (так что здесь $b>a$), то

$$\int\limits_a^b v(t)\,dt<0.$$

В случае движения смысл ответа прост: отрицательное значение v означает, что движение происходит в сторону, противоположную положительному направлению, т. е. направлению возрастания координаты z. Путь, пройденный в отрицательном направлении, мы всегда считаем отрицательным. При таком движении z уменьшается, $z(b)<z(a)$. Общая формула

$$z(b)=z(a)+z(a,\ b)=z(a)+\int\limits_a^b v\,dt$$

и $\int\limits_a^b v(t)\,dt=z(b)-z(a)$

остается при этом в силе — только во второй ее форме обе части равенства отрицательны.

В случае знакопеременной скорости может случиться, в частности, что $\int\limits_a^b v(t)\,dt=0$, хотя $b>a$ и $b\neq a$. Это произойдет, если часть времени от a до b тело двигалось в одну сторону, а другую часть времени — в противоположную сторону, в результате чего к моменту b оно вернулось в исходное положение, т. е. в то положение, в котором находилось в момент a.

Обратимся к задаче о площади под кривой. При $b>a$ и $v(t)>0$ интеграл равен площади, ограниченной кривой $v=v(t)$, осью t и вертикалями $t=a$, $t=b$ (рис. 1, *а*). При $v<0$ и $b>a$, как мы знаем, $\int\limits_a^b v\,dt<0$. В этом случае кривая лежит ниже оси абсцисс (рис. 1, *б*); поэтому, для того чтобы сохранить закон, по которому площадь равна интегралу, необходимо считать площадь отрицательной,

когда кривая $v=v(t)$ лежит ниже оси абсцисс.

Возьмем теперь знакопеременную функцию, например синусоиду $v(t)=$

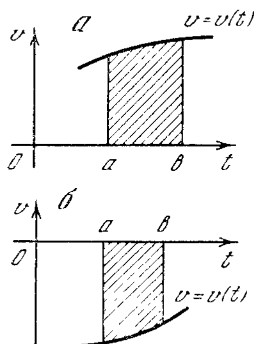


Рис. 3.5.1

$=\sin t$ (рис. 2). Нетрудно видеть, что площадь, задаваемая этой кривой на отрезке, равном периоду функции, т. е. от $t=0$ до $t=2\pi$, по нашему определению, равна нулю. Это значит, что площадь первой полуволны, которую мы считаем положительной, сокращается с отрицательной площадью второй полуволны [1].

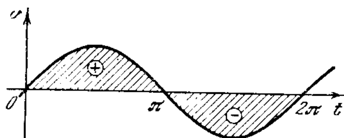


Рис. 3.5.2

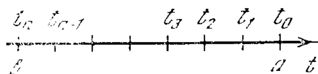


Рис. 3.5.3

Определенный интеграл обобщается и на случай, когда верхний предел меньше нижнего. В этом случае мы уже не будем говорить о пути, времени, скорости (ср. § 1), а обратимся к определению интеграла как суммы (см. § 2). Разбивая снова отрезок от a до b (рис. 3) на n частей промежуточными значениями $t_1, t_2, \ldots, t_{n-1}$, убедимся, что все Δt теперь отрицательны. Легко проверить, что

$$\int_a^b v(t)\,dt = -\int_b^a v(t)\,dt, \tag{3.5.1}$$

так как при любом разбиении отрезка $[a, b]$ соответствующие суммы будут отличаться знаками всех Δt во всех слагаемых.

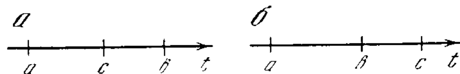


Рис. 3.5.4

Существенное свойство интеграла состоит в том, что область интегрирования можно разбить на части: путь, пройденный за время от a (начало) до b (конец), можно представить как сумму пути, пройденного за время от a до c (промежуточного момента времени) и от c до b (рис. 4, *а*):

$$\int_a^b v(t)\,dt = \int_a^c v(t)\,dt + \int_c^b v(t)\,dt. \tag{3.5.2}$$

При помощи соотношения (1) можно распространить формулу (2) и на случай, когда c лежит не внутри промежутка $[a, b]$, а вне его. Пусть, например, $c > b > a$ (рис. 4, *б*). Тогда, очевидно,

$$\int_a^c v(t)\,dt = \int_a^b v(t)\,dt + \int_b^c v(t)\,dt. \tag{3.5.3}$$

Перенесем последнее слагаемое влево и воспользуемся (1):

$$\int_a^c v\,dt - \int_b^c v\,dt = \int_a^c v(t)\,dt + \int_c^b v(t)\,dt = \int_a^b v\,dt. \tag{3.5.4}$$

Таким образом, мы получили равенство, в точности совпадающее с (2).

Аналогично можно рассмотреть все другие случаи взаимного расположения чисел a, b, c — точек числовой оси (всего имеется шесть вариантов расположения этих точек). Читатель легко сможет рассмотреть их самостоятельно и убедиться, что формула (2) верна

[1] Если перед нами стоит вопрос о том, сколько нужно краски, чтобы закрасить заштрихованные на рис. 2 области, то такое определение площади, разумеется, будет бесполезно. В этом случае надо разбить весь интеграл на части, в каждой из которых $\sin x$ сохраняет один и тот же знак (в данном случае на две части — от 0 до π и от π до 2π), подсчитать интеграл по каждой части отдельно и сложить абсолютные величины интегралов, относящихся к отдельным частям.

во всех этих случаях, т. е. независимо от взаимного расположения чисел (точек) a, b, c.

Все эти свойства определенных интегралов мы вывели исходя из определения интеграла как предела суммы. Но эти свойства следуют также из выражения определенного интеграла через неопределенный. В самом деле, пусть неопределенный интеграл

$$\int v(t)\,dt = z(t).$$

Тогда

$$\int_a^b v(t)\,dt = z(b) - z(a)$$

и

$$\int_b^a v(t)\,dt = z(a) - z(b) = -\int_a^b v(t)\,dt.$$

Аналогично

$$\int_a^b v(t)\,dt = z(b) - z(a), \quad \int_a^c v(t)\,dt = z(c) - z(a) \quad \text{и} \quad \int_c^b v(t)\,dt = z(b) - z(c),$$

откуда

$$\int_a^b v(t)\,dt = \int_a^c v(t)\,dt + \int_c^b v(t)\,dt.$$

Основной закон о том, что производная от интеграла равна подынтегральной функции, относится к производной по верхнему пределу. Если определенный интеграл рассматривать как функцию его нижнего предела при закрепленном (постоянном) верхнем пределе, то ответ получит знак «минус»:

$$\frac{dz(a,\ b)}{da} = \frac{d}{da}\left(\int_a^b v(t)\,dt\right) = -v(a). \qquad (3.5.5)$$

Причину появления знака минус в этой формуле легко понять, рассматривая интеграл как площадь: положительное приращение Δa, очевидно, уменьшает площадь (рис. 5)[2]. Формально тот же результат можно получить, переставив пределы (при этом нам придется изменить знак) и затем применяя известный закон о производной по верхнему пределу:

$$\frac{d}{da}\left(\int_a^b v(t)\,dt\right) = \frac{d}{da}\left(-\int_b^a v(t)\,dt\right) = -v(a).$$

В связи с вопросом о знаке интеграла отметим пример, часто вызывающий недоумение у начинающих. Рассмотрим

$$\int \frac{dx}{x^2} = -\frac{1}{x}. \qquad (3.5.6)$$

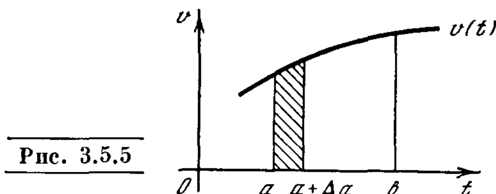


Рис. 3.5.5

Это равенство вытекает из найденного ранее значения производной

$$\frac{d(1/x)}{dx} = -\frac{1}{x^2}.$$

Правилен ли знак у интеграла? Может ли быть отрицательным интеграл от положительной функции $\frac{1}{x^2}$? Не противоречит ли этот знак сделанным выше утверждениям?

Наше недоумение связано с тем, что формулу (6) мы записали неаккуратно. Если переписать ее в полном виде

$$\int \frac{dx}{x^2} = -\frac{1}{x} + C,$$

то уже нельзя будет говорить о том, что знак интеграла всегда отрицателен, так как это зависит еще от знака и величины C.

В действительности все утверждения о знаке относились к о п р е д е л е н н о м у интегралу. Возьмем

$$\int_a^b \frac{1}{x^2}\,dx = \left(-\frac{1}{x}\right)\Big|_a^b = \left(-\frac{1}{b}\right) - \left(-\frac{1}{a}\right) = \frac{1}{a} - \frac{1}{b} = \frac{(b-a)}{ab}.$$

При $b > a$ (и если притом и a и b, каждое в отдельности, положительны) интеграл положителен, как и должно быть, т. е. формула (6) правильна,

[2] Площадь, ограниченная вертикальными прямыми $t=a+\Delta a$, $t=b$, кривой $v=v(t)$ и осью t, м е н ь ш е площади, ограниченной вертикальными прямыми $t=a$, $t=b$, кривой и осью t.

приводит к верному результату для определенного интеграла. Забегая вперед, отметим, что с интегралом $\int \frac{dx}{x^2}$ связаны другие, уже не фиктивные, а истинные трудности, которые будут рассмотрены в § 5.2.

§ 6. Примеры и приложения

В гл. 2 и предыдущих параграфах гл. 3 мы рассматривали соотношение между путем и скоростью, соотношение между уравнением кривой и площадью под этой кривой. Эти соотношения представляют собой те конкретные вопросы, на почве которых исторически сложились дифференциальное и интегральное исчисления. Но понятия производной и интеграла применимы, конечно, не только к перечисленным вопросам, а к чрезвычайно широкому кругу явлений, к самым различным областям науки, техники, жизни. В сущности говоря, производная, интеграл и теорема Ньютона—Лейбница, фиксирующая связь между ними, представляют собой определенный язык, наиболее приспособленный для описания законов природы.

Учащийся, начинающий изучение иностранного языка, чтобы привыкнуть к нему, повторяет однотипные простые фразы: «на столе стоит стакан», «на столе лежит хлеб», «на полу сидит кошка», «на полу сидит мышь». Вот так же в начале изучения высшей математики надо на многих похожих примерах повторять соотношения между производной и интегралом. Сперва надо научиться иностранному языку, а потом уже высказывать на этом языке определенные мысли, утверждения, желания. Так и мы сначала научимся выражать хорошо известные соотношения и формулировать задачи на языке математического анализа, а уже потом будем решать эти задачи и получать новые результаты [1].

Приведем несколько типичных примеров, иллюстрирующих содержание настоящей главы; большинство этих примеров будут продолжены и развиты в дальнейших главах.

1. Ускорение и равноускоренное движение; работа силы. В гл. 2 и 3 мы постоянно рассматривали перемещение тел и скорость движения как производную от координаты тела по времени. Но после того как мы нашли мгновенную скорость $v(t)=\frac{dz(t)}{dt}$ и определили ее зависимость от времени, становится уместным вопрос о «скорости изменения скорости». Бегло этого вопроса мы уже касались в § 2.7.

Производная скорости $v(t)$ по времени

$$\frac{dv(t)}{dt}=a(t) \qquad (3.6.1)$$

называется *ускорением* и обозначается обычно буквой a (accélération — ускорение по-французски). Так как размерность скорости см/с или м/с, то размерность ускорения см/с^2 или м/с^2.

Запишем скорость в виде производной от пути по времени $v=\frac{dz}{dt}$ и подставим в (1). Тогда получим

$$a=\frac{d}{dt}\left(\frac{dz}{dt}\right)=\frac{d^2z}{dt^2}. \qquad (3.6.2)$$

[1] Гете говорил: «Математики — это род французов: стоит им что-либо сказать, как они сразу переводят сказанное на свой язык — и это оказывается совсем не тем, что Вы первоначально имели в виду». С этим перекликается известное высказывание замечательного американского физика, творца современной термодинамики Джозайи Уилларда Гиббса (1839—1903). При обсуждении вопроса о том, чему следует в программах оказывать предпочтение: языкам, особенно латыни и греческому, или математике, Гиббс сказал: «Математика — это тоже язык». Это было его единственное — и тем более весомое — публичное выступление по общим вопросам преподавания. Еще много раньше почти то же утверждал знаменитый Галилео Галилей (1564—1642), заявивший, что «законы природы записаны в величайшей книге, которая всегда открыта перед нашими глазами (я разумею Вселенную), но ее нельзя понять, не научившись сначала понимать язык, на котором она написана, — а написана она на языке математики». (Правда, далее Галилей несколько неожиданно для нас с Вами заявляет, что «ее (математики) буквы — это треугольники, окружности и другие геометрические фигуры, без которых невозможно понять ее слова», рассматривая тем самым как язык природы геометрию, а не дифференциальное и интегральное исчисления. Но ведь надо иметь в виду, что во времена Галилея никакой математики, отличной от геометрии древних греков, еще не было — предвидеть создание Лейбницем и Ньютоном высшей математики их прямой предшественник Галилей, разумеется, не мог.)

Таким образом, *ускорение равно второй производной от пути по времени.*

Заметим, как разумно поставлены показатели (двойки) в выражении $\frac{d^2z}{dt^2}$. Размерность ускорения именно $\frac{z}{t^2}$; отбрасывая безразмерные знаки d, получаем правильную размерность второй производной.

Если ускорение как функция времени нам известно, то значение $v = v(t)$ мгновенной скорости можно записать в виде интеграла от $a(t) = \frac{dv}{dt}$:

$$v(t) = v_0 + \int_{t_0}^{t} a(t)\,dt. \qquad (3.6.3)$$

(Здесь $v_0 = v(t_0)$ — мгновенная скорость в начальный момент времени $t = t_0$; чтобы убедиться в этом, достаточно подставить в обе части (3) значение $t = t_0$.) В частности, если ускорение постоянно (a = const, случай *равноускоренного* движения), то из (3) следует

$$v(t) = v_0 + a\int_{t_0}^{t} dt = v_0 + a(t - t_0), \qquad (3.6.4)$$

или

$$v(t) = at + b, \text{ где } b = v_0 - at_0, \qquad (3.6.4a)$$

т. е. скорость равноускоренного движения является линейной функцией времени.

Зная скорость $v = v(t)$ движения, мы можем вычислить и координату $z = z(t)$ (т. е. выписать *закон движения*):

$$z(t) = z_0 + \int_{t_0}^{t} v(t)\,dt, \qquad (3.6.5)$$

или в силу (3)

$$z(t) = z_0 + \int_{t_0}^{t}\left[v_0 + \int_{t_0}^{t} a(t)\,dt\right]dt =$$

$$= z_0 + v_0(t - t_0) + \int_{t_0}^{t}\left[\int_{t_0}^{t} a(t)\,dt\right]dt, \qquad (3.6.5a)$$

где $z_0 = z(t_0)$ — начальное положение (координата) тела в момент t_0 (подставьте в обе части (5) значение $t = t_0$). Разумеется, в общем своем виде формула (5a) малонаглядна. Однако если a=const, то скорость v задается несложной формулой (4a) (или (4)) и соотношение (5a) позволяет полностью описать закон движения тела:

$$z(t) = z_0 + \int_{t_0}^{t}(at + b)\,dt =$$

$$= z_0 + a\frac{t^2 - t_0^2}{2} + b(t - t_0) = \frac{a}{2}t^2 + bt + c,$$

$$\text{где } c = -\frac{a}{2}t_0^2 - bt_0 + z_0 \qquad (3.6.6)$$

(проверьте, что производная от правой части (6) по t совпадает с правой частью (4a)). Таким образом, зависимость координаты z тела от времени t при равноускоренном движении является квадратичной. В частности, в случае свободного падения тел имеем $a = -g$, где $g \approx 9{,}8$ м/с² (знак минус связан с тем, что положительным мы считаем направление вверх). Полагая в (4), (4a) и (6) $a = -g$, получим известные формулы:

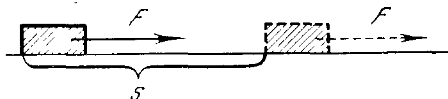


Рис. 3.6.1

$$v(t) = v_0 - \int_{t_0}^{t} g\,dt = v_0 - g(t - t_0) =$$

$$= v_0 - gt_1,$$

где $t_1 = t - t_0$ — это «новое» время, отсчитываемое от момента $t = t_0$, и

$$z(t) = z_0 + \int_{t_0}^{t}[v_0 - g(t - t_0)]\,dt =$$

$$= z_0 + \int_{0}^{t_1}(v_0 - gt_1)\,dt_1 = z_0 + v_0t_1 - \frac{g}{2}t_1^2.$$

В связи с ускорением вспомним еще, что *работа* силы равна произведению силы F на путь s (мы здесь считаем путь или перемещение совпадающим с направлением силы; рис. 1). Однако эта формула применима лишь в случае постоянства силы — обстоятельство, которое реально почти никогда места не имеет. В случае же переменной силы F приходится разбивать весь путь (от $z = z_0$ до какого-то другого (переменного) значения координаты z) на малые интервалы длины Δz и суммировать выражения $F\Delta z$, получаемые в пред-

положении, что на рассматриваемом малом интервале длины Δz сила не меняется (еще не успела ощутимо измениться). Так мы приходим к «интегральной сумме» $\Sigma F\Delta z$, в которой для получения выражения для проделанной работы A надо еще затем перейти к пределу, считая все отрезки Δz неограниченно убывающими. Последнее обстоятельство позволяет считать скорость $v=\frac{dz}{dt}$ на каждом таком отрезке постоянной и потому выразить длину Δz отрезка как произведение скорости v на время Δt, в течение которого этот отрезок проходится. Тем самым сумма выражений $\sum F\Delta z$ заменяется на сумму $\sum Fv\Delta t$, что позволяет представить работу A либо в виде интеграла, где за переменную принята координата z, либо в виде интеграла, где переменной служит время t:

$$A=\int_{z_0}^{z} Fdz=\int_{t_0}^{t} Fvdt. \qquad (3.6.7)$$

Вспомним теперь, что сила F равна произведению ma, где m — масса тела, а $a=\frac{dv}{dt}$—ускорение (закон Ньютона!). Поэтому (7) можно переписать так:

$$A=m\int_{t_0}^{t} \frac{dv}{dt}vdt. \qquad (3.6.8)$$

Поскольку, очевидно, функция $v\frac{dv}{dt}$ является производной от $\frac{v^2}{2}$ (ср. § 2.3), то окончательно получаем

$$A=m\left(\frac{v^2}{2}\right)\Big|_{t_0}^{t}=\frac{mv^2}{2}-\frac{mv_0^2}{2}, \qquad (3.6.9)$$

где $v=v(z)$ и $v_0=v(z_0)$: проделанная *работа* A равна *приращению* $\frac{mv^2}{2}-\frac{mv_0^2}{2}$ *кинетической энергии* $\frac{mv^2}{2}$ *тела* (разумеется, работа A и приращение кинетической энергии могут одновременно оказаться отрицательными).

По поводу дальнейшего обсуждения затронутых здесь тем см. главы 9 и 10 книги.

2. Растяжение проволоки. Рассмотрим подвешенный за верхний конец прут, представляющий собой метровый кусок медной проволоки диаметром 0,4 см; спрашивается, до какой длины растянется этот кусок проволоки под влиянием собственного веса? Растяжение проволоки связано с приложенной силой с помощью так называемого закона Гука[2], справедливого при сравнительно небольших силах (и растяжениях) и утверждающего, что *деформация* (*растяжение*) l *прямо пропорциональна приложенной силе* F *и длине* L *проволоки и обратно пропорциональна площади* S *ее сечения*:

$$l=\frac{1}{E}\frac{FL}{S}, \qquad (3.6.10)$$

где E — постоянный множитель, зависящий от материала, из которого сделана проволока, и называемый модулем Юнга.

Разумеется, мы не можем сразу подставить в формулу (10) вместо L длину 1 м (100 см) куска проволоки, а вместо F — силу тяжести, легко находимую по объему прута и плотности меди, ибо действующая на отдельные участки подвешенной проволоки сила будет разной для разных участков: если для точки подвеса A эта сила действительно будет равна весу всего куска проволоки, то для второго (свободного) конца B она будет равна нулю, ибо этот конец совсем не будет нагружен (рис. 2). Поэтому нам приходится разбить всю проволоку, которой отвечает фиксирующая высоты координата z, на малые участки длины Δz: для каждого такого участка, удаленного на z от свободного конца B проволоки, силу можно считать постоянной и равной весу части прута длины z: $F=\rho gSz$, где ρ — плотность меди, g — ускорение свободного падения, $V=Sz$— объем рассматриваемой части прута. Таким образом, этому малому участку проволоки длины Δz отвечает удлинение

$$\Delta l=\frac{\rho g}{E}\frac{Sz\Delta z}{S}=\frac{\rho g}{E}z\Delta z, \qquad (3.6.11)$$

а общее удлинение всей проволоки получается суммированием (интегрированием) всех таких «элементарных удлинений»:

$$l=\frac{\rho g}{E}\int_0^{100} zdz=\frac{\rho g}{E}\left(\frac{z^2}{2}\right)\Big|_0^{100}=\frac{\rho g}{E}\frac{(100)^2}{2}. \qquad (3.6.11a)$$

[2] Роберт Гук (1635—1703) — выдающийся английский ученый, современник (и во многих случаях научный оппонент) И. Ньютона.

Подставляя сюда численные значения $g=980$ см/с^2, $\rho=8,9$ г/см^3, $E=9,8\cdot10^{13}$ г/(см·с^2), получаем

$$l=\frac{980\cdot8,9\cdot10^4}{2\cdot9,8\cdot10^{13}}\approx4,5\cdot10^{-5}\text{ см}$$

(заметьте, что множитель $(100)^2=10^4$ в числителе дроби имеет размерность см^2).

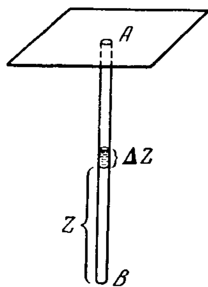
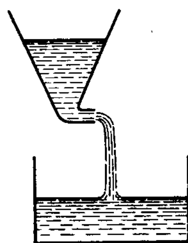


Рис. 3.6.2


Рис. 3.6.3

3. Вытекание жидкости. Представим себе сосуд произвольной формы (рис. 3), из которого вытекает жидкость. Масса жидкости, находящейся в данный момент в сосуде, равна M. Эта величина есть функция времени: $M=M(t)$. Жидкость собирается в другом сосуде; массу ее во втором сосуде обозначим через $m(t)$. Пусть $W=W(t)$ — масса жидкости, вытекающая из первого сосуда в единицу времени; эта величина — *поток* жидкости — имеет размерность г/с. Величины m, M и W связаны между собой соотношениями

$$\frac{dM}{dt}=-W(t),\quad \frac{dm}{dt}=W(t). \tag{3.6.12}$$

Те же (дифференциальные, т. е. связанные с производными) соотношения можно записать в интегральном виде. Пусть в некоторый начальный момент времени t_0 масса жидкости в первом сосуде равнялась $M(t_0)=M_0$, а второй сосуд был пуст: $m(t_0)=0$. Тогда

$$m(t_1)=\int_{t_0}^{t_1}W(t)\,dt,$$

$$M(t_1)=M(t_0)-\int_{t_0}^{t_1}W(t)\,dt. \tag{3.6.13}$$

Таким образом, масса жидкости в определенный момент времени t_1 выражается интегралом, в котором переменная интегрирования пробегает все значения от t_0 до t_1.

Удобнее заменить в (13) t_1 на t, а переменную интегрирования переименовать (воспользовавшись тем, что она немая), обозначив ее, скажем, буквой τ. Тогда получим

$$m(t)=\int_{t_0}^{t}W(\tau)\,d\tau,$$

$$M(t)=M(t_0)-\int_{t_0}^{t}W(\tau)\,d\tau. \tag{3.6.13а}$$

Впрочем, чаще просто пишут

$$m(t)=\int_{t_0}^{t}W(t)\,dt,$$

$$M(t)=M(t_0)-\int_{t_0}^{t}W(t)\,dt, \tag{3.6.13б}$$

не обращая внимания на то, что здесь одна буква t имеет два разных смысла: переменной интегрирования и аргумента функций $M(t)$ и $m(t)$ (совпадающего с верхним пределом интегралов). В этом отношении записи (13) и (13а) точнее, чем (13б).

Написанные выше формулы соответствуют опыту, в котором в различные моменты времени измеряются M и поток жидкости W.

Но часто задача ставится так: W — расход жидкости зависит известным образом от ее давления, т. е. от высоты столба жидкости h. В свою очередь, при данной форме сосуда величина h зависит от M. Таким образом, известен расход W как функция от массы жидкости, находящейся в сосуде: $W=W(M)$. Тогда равенство (12) приобретает вид

$$\frac{dM}{dt}=-W(M). \tag{3.6.14}$$

Равенства такого рода, связывающие неизвестную функцию (в нашем случае — функцию $M=M(t)$) с ее производной, называются *дифференциальными уравнениями*; таким образом (14) — это дифференциальное уравнение с неизвестной (подлежащей определению) функцией $M(t)$. Тему о дифференциальных уравнениях, а также обсуждение поставленного здесь вопроса мы продолжим в гл. 6.

4. Конденсатор. Накопленный в конденсаторе К (рис. 4) заряд

(количество электричества) обозначим через q. В системе единиц СИ q измеряется в кулонах (обозначение — Кл). Электрический ток j, протекающий по проводу, представляет собой количество электричества, протекающее в единицу времени; он измеряется в амперах (обозначение — А). Один ампер есть ток в один кулон в секунду: $1\,А=1$ Кл/с; напротив, количество электричества имеет размерность А·с (в си-

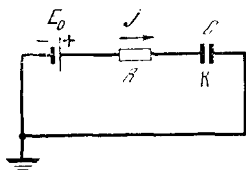


Рис. 3.6.4

стеме СИ единица А относится к числу основных).

Заряд конденсатора [3] и ток связаны между собой равенством

$$\frac{dq}{dt}=j \qquad (3.6.15)$$

(положительное направление тока на рис. 4 указано стрелкой). Если задана или экспериментально определена зависимость $j=j(t)$ тока от времени, то можно написать интегральное соотношение

$$q(t_1)=q(t_0)+\int_{t_0}^{t_1} j(t)\,dt. \qquad (3.6.15а)$$

Если дана *емкость* C конденсатора, то падение напряжения на конденсаторе можно выразить через q:

$$\varphi_C=\frac{q}{C},$$

тогда падение напряжения на сопротивлении R равно

$$\varphi_R=E_0-\varphi_C=E_0-\frac{q}{C},$$

где E_0 — напряжение батареи. По закону Ома текущий через сопротивление R ток равен $j=\frac{1}{R}\left(E_0-\frac{q}{C}\right)$, и, пользуясь (15), получаем д и ф ф е р е н ц и а л ь н о е у р а в н е н и е

$$\frac{dq}{dt}=\frac{1}{R}\left(E_0-\frac{q}{C}\right). \qquad (3.6.16)$$

[3] *Зарядом конденсатора* мы будем называть выраженное в кулонах количество положительного электричества на левой пластине конденсатора К (см. рис. 4).

Подробно задачи с конденсаторами будут рассмотрены в последней главе второй части книги.

5. А т м о с ф е р н о е д а в л е н и е. Мысленно выделим в атмосфере вертикальный столб воздуха с постоянным сечением S (см^2). Плотность воздуха ρ (г/см^3) зависит от высоты h над поверхностью Земли. Объем тонкого слоя, заключенного между h и $h+\Delta h$ (рис. 5), равен $S\Delta h$. Внутри этого тонкого слоя плотность $\rho(h)$ можно считать неизменной — именно для этого слой и брался тонким. В данном примере под Δh можно понимать величину 1 или 10 м и даже (с несколько меньшей точностью) 100 м, поскольку при изменении высоты на 1 км плотность воздуха меняется всего на 12—14 %.

Объем слоя толщины Δh равен $S\Delta h$, а, значит, масса воздуха в этом слое, по определению плотности, равна $\Delta m=\rho S\Delta h$.

Для того чтобы найти массу воздуха в столбе, простирающемся от высоты h_1 до h_2, нам надо составить сумму выражений $\rho S\Delta h$, где S постоянно, а ρ меняется с изменением h. Эта сумма распространена на все части Δh, на которые разбит наш столб воздуха. Неограниченно уменьшая все толщины

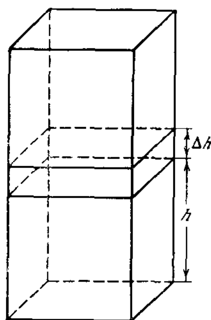


Рис. 3.6.5

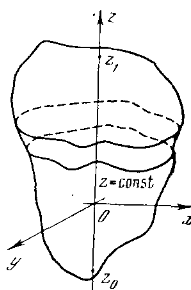


Рис. 3.6.6

Δh отдельных слоев, мы приходим к интегралу

$$m(h_1,\ h_2)=S\int_{h_1}^{h_2}\rho(h)\,dh. \qquad (3.6.17)$$

Масса воздуха в столбе от поверхности Земли ($h=0$) до высоты h равна

$$m(0,\ h)=S\int_0^h \rho(h)\,dh. \qquad (3.6.17а)$$

Масса воздуха, находящегося выше заданной высоты h, равна

$$m = S\int_h^\infty \rho(h)\,dh, \tag{3.6.17б}$$

где символ ∞ в качестве верхнего предела интеграла заменяет очень большое число H — такое большое, что при дальнейшем его увеличении величина интеграла практически не меняется.

Давление P на некоторой высоте h, умноженное на площадь S, равно силе, с которой притягивается к Земле весь столб воздуха, находящийся выше высоты h. Сила тяготения равна массе, умноженной на ускорение свободного падения[4] g, откуда

$$P(h) = \int_h^\infty g\rho(h)\,dh. \tag{3.6.18}$$

Пользуясь формулой (5.5), получим отсюда д и ф ф е р е н ц и а л ь н о е у р а в н е н и е

$$\frac{dP}{dh} = -g\rho(h). \tag{3.6.19}$$

Уравнение (19) можно было бы написать и сразу, рассматривая равновесие тонкого слоя dh, на который снизу действует давление $P(h)$, сверху давление $P(h+dh)$; равнодействующая этих двух сил уравновешивает притяжение к Земле массы, заключенной в слое.

Далее эта тема будет продолжена в гл. 11.

6. О б ъ е м. Используем интегральное исчисление для вычисления объемов тел. Разобьем тело общего объема V (рис. 6) на тонкие пластинки плоскостями $z=\text{const}$; здесь мы используем систему координат в пространстве, задаваемую осью абсцисс x, осью ординат y на (горизонтальной) плоскости xOy, а также (вертикальной) *осью аппликат* Oz (третья координатная ось в пространстве). Объем ΔV тонкого слоя тела, ограниченного плоскостями, соответствующими значениям z и $z+\Delta z$ аппликаты, приближенно равен произведению площади $S(z)$ сечения на толщину Δz слоя. Отсюда следует, что если площадь $S=S(z)$ горизонтального сечения тела нам известна, то объем тела может быть вычислен по формуле

$$V = \int_{z_0}^{z_1} S(z)\,dz, \tag{3.6.20}$$

где z_0 и z_1 — наименьшее и наибольшее значения аппликаты в пределах нашего тела.

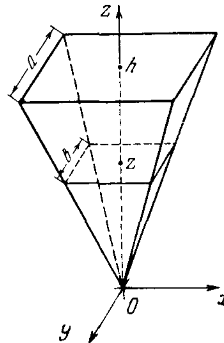


Рис. 3.6.7

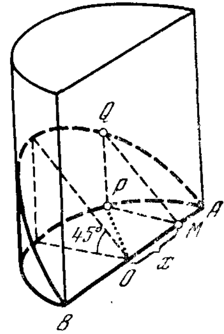


Рис. 3.6.8

Применим эту формулу к правильной четырехугольной пирамиде. Поставим пирамиду на ее вершину, причем совместим вершину с началом координат, а ось симметрии пирамиды направим по оси z (рис. 7). Пусть высота пирамиды равна h, а основание ее (оказавшееся сверху) представляет собой квадрат со стороной a. Из геометрии известно, что сечение пирамиды горизонтальной плоскостью на высоте z представляет собой также квадрат, сторона b которого относится к a, как z к h: $b = b(z) = a\frac{z}{h}$. Следовательно, площадь сечения $S(z) = b^2 = \frac{a^2}{h^2}z^2$. Объем пирамиды

$$V = \int_0^h \frac{a^2}{h^2}z^2dz = \frac{a^2}{h^2}\int_0^h z^2dz.$$

Воспользуемся результатом § 4:

$$\int z^2dz = \frac{1}{3}z^3, \quad \int_0^h z^2dz = \frac{1}{3}h^3.$$

[4] В большинстве случаев можно считать допустимым предположение о постоянстве для рассматриваемых высот h над уровнем Земли ускорения свободного падения g и поверхность Земли предполагать плоской; впрочем, оба эти предположения никак не отражаются на наших рассуждениях.

Тогда получим выражение объема пирамиды

$$V=\frac{a^2}{h^2}\cdot\frac{1}{3}h^3=\frac{1}{3}a^2h.$$

Объем пирамиды равен одной трети произведения площади основания на высоту пирамиды. Выводы этой формулы в стереометрии без применения интегралов значительно сложнее.

К вопросу о использовании интегрального исчисления для вычисления объемов мы вернемся в гл. 7.

Вот еще один пример, гораздо более эффектный, чем разобранные выше. Поставим задачу определения объема *цилиндрического «копыта»*, т. е. части (прямого кругового) цилиндра, отсекаемого от него проходящей через диаметр AB основания плоскостью, образующей с основанием угол 45° (рис. 8); радиус цилиндра для простоты примем равным 1 (ср., впрочем, упр. 3). Сечение копыта плоскостью, перпендикулярной диаметру AB и удаленной от центра O основания на расстояние $OM=x$, представляет собой прямоугольный равнобедренный треугольник (прямоугольный треугольник MPQ с острым углом $\widehat{QMP}=45°$), катет MP которого равен $\sqrt{OP^2-OM^2}=\sqrt{1-x^2}$. Но тогда и $PQ=PM=\sqrt{1-x^2}$ и $S_{\triangle MPQ}=S(x)=\frac{1}{2}MP\cdot PQ=\frac{1}{2}(1-x^2)$. Поэтому в силу формулы (20), где теперь надо только писать x вместо z, объем V копыта равен

$$V=\int_{-1}^{1}\frac{1}{2}(1-x^2)\,dx=\frac{1}{2}\int_{-1}^{1}(1-x^2)\,dx=$$
$$=\frac{1}{2}\left[\int_{-1}^{1}dx-\int_{-1}^{1}x^2dx\right]=\frac{1}{2}\left[x\Big|_{-1}^{+1}-\frac{x^3}{3}\Big|_{-1}^{+1}\right]=$$
$$=\frac{1}{2}\left[(1-(-1))-\left(\frac{1}{3}-\frac{-1}{3}\right)\right]=$$
$$=\frac{1}{2}\left(2-\frac{2}{3}\right)=\frac{2}{3}$$

(здесь мы используем очевидные свойства интеграла: $\int_a^b cf\,dx=c\int_a^b f\,dx$, где c — число, и $\int_a^b (f+g)\,dx=\int_a^b f\,dx+\int_a^b g\,dx$).

Заметьте, что объем рассматриваемого «круглого тела» равен рациональной дроби и никак не связан с числом π!

Упражнения

3.6.1. Выведите формулу для объема (произвольной) пирамиды, используя свойства параллельных сечений.

3.6.2. Найдите объем (прямого кругового) конуса высоты h, основанием которого служит круг радиуса R. Изменится ли результат, если рассматриваемый (круговой) конус не будет прямым, т. е. если его вершина не будет расположена в точности над центром основания?

3.6.3. Найдите объем «копыта», отсекаемого от цилиндра радиуса R плоскостью, проходящей через диаметр основания цилиндра и составляющей с основанием известный угол α.

3.6.4. Воспользовавшись формулой для объема цилиндрического копыта, найдите величину интеграла $\int_0^1 y\sqrt{1-y^2}\,dy$.

* * *

В гл. 2 и 3 рассмотрены понятия производной и интеграла, некоторые самые простые их свойства и раскрыта связь между интегралом и производной — другими словами, здесь намечен основной костяк понятий и теорем так называемой высшей математики. Техника дифференцирования и интегрирования, т. е. вопросы практического вычисления производных и интегралов от различных функций, отнесены к гл. 4—5 книги; в гл. 2 и 3 приведено только несколько самых простых примеров.

Хочется посоветовать учащемуся не измерять важность (и даже трудность) того или иного раздела числом формул, их сложностью и громоздкостью. В действительности самое важное и трудное — именно математическая формулировка задачи в виде алгебраического уравнения, или интеграла, или дифференциального уравнения, но не следующие за этой стадией преобразования. Именно на это надо обратить главное внимание.

Почти каждый физик знает, что те работы, которые ему не удалось сделать (и которые тем временем были сделаны другими!), не были им сделаны потому, что, ограничиваясь общим размышлением, он не решался перевести свои мысли на математический

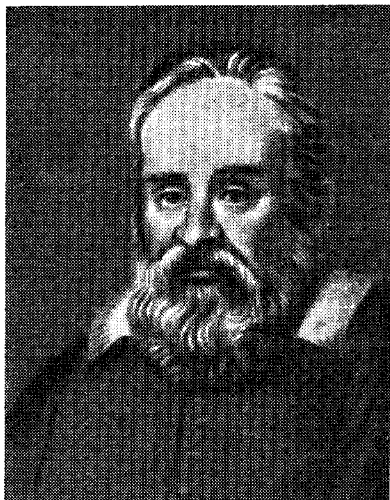
язык, облечь их в форму уравнений. Что же до вычислительных трудностей в четко поставленной задаче с ясным физическим содержанием, то они всегда преодолеваются — если не точным расчетом, то приближенными методами.

Если последние три параграфа показались читателю трудными, целесообразно перечитать гл. 2 и 3 еще раз.

Возникновение «высшей математики», т. е. *дифференциального и интегрального исчисления*, явилось переломным моментом во всей истории человеческой культуры: с их созданием люди получили в руки мощный аппарат, приспособленный для анализа всевозможных процессов, для глубокого объяснения физических явлений, для построения научной картины мира. Собственно, первые задачи, относящиеся к дифференциальному и интегральному исчислениям, с большим искусством решали еще древнегреческие мыслители, в первую очередь гениальный Архимед Сиракузский (287—212 до н. э.), успешно применявший математику к конструированию механизмов и машин (созданная им военная техника привела в ужас римлян, осаждавших его родные Сиракузы). Так, Архимед умел проводить касательные к кривым и определять площади, ограниченные кривыми линиями: он успешно решил задачу проведения касательной к так называемой *архимедовой спирали*, т. е. к линии, которую описывает улитка, равномерно ползущая по спице колеса, вращающегося с постоянной угловой скоростью, а также задачу *квадратуры параболы*, т. е. нахождения площади сегмента параболы. Однако в то время подобные задачи, которые ныне без всякого труда решает любой студент или даже школьник-старшеклассник, были доступны разве что Архимеду: общего метода, применимого ко всем задачам такого рода, не существовало, и каждая из них требовала весьма значительных усилий. Да примитивный уровень античной техники и не требовал решения таких задач, которые лишь свидетельствовали об изобретательности решившего их ученого, но не применялись ни в каких внематематических проблемах: мыслители древности, как правило, рассматривали математику не как мощный метод решения практических задач, а лишь как теоретическую науку, совершенство которой копирует глубокую гармонию мира, но объясняет мир лишь в чисто философском смысле. (Чуть ли не единственным исключением здесь был тот же Архимед, в трудах которого математика органически переплеталась с механикой и физикой. Но ведь Архимед был гений, — и здесь, как и во многом другом, он значительно обогнал свое время.) Статический характер жизни в Древней Греции, почти не знавшей серьезных механизмов, жизни, концентрирующейся на городских площадях, вокруг дворцов и храмов, украшенных прекрасными статуями, столь же неподвижными, как сами храмы или городские площади, породил свойственную математике древних *метафизику* — запрещение рассматривать текущие процессы и стремление ограничиться одними лишь застывшими «состояниями», отражением которых для знаменитого Евклида Александрийского (начало 3 в. до н. э.) явились столь частые в его рассуждениях цепочки равных (конгруэнтных) треугольников.

Метафизика древних стала немыслима в условиях расцвета итальянских городов в XV—XVI вв. и появления первых мануфактур, предвещавших возникновение в скором будущем машинного производства, — в этот период она могла лишь тормозить необходимый научный прогресс. Великий Галилео Галилей (1564—1642) впервые громко провозгласил, что ключом к тайнам Вселенной является математика (см. сноску на с. 98). Под влиянием Галилея его непосредственные ученики — создатель барометра Эванджелиста Торричелли (1608—1647) и геометр Бонавентура Кавальери (1598—1647), продолжая дело столь любимого их учителем Архимеда, решили много частных задач, которые сегодня относят к высшей математике. В частности, Кавальери принадлежит метод вычисления объемов, очень близкий к развитым выше построениям. Галилей же впервые заметил, что задача *об определении пути по известной скорости* практически совпадает со столь занимавшей еще Архимеда задачей *о нахождении площадей криволинейных фигур*; продолжая Галилея, Торричелли установил, что обратная задача *об отыскании скорости по пути* родственна задаче *проведения касательной к (кривой) линии*. Но общих методов решения таких задач, общего алгоритма, позволяющего «считать, не рассуждая», в то время еще не было.

Не было подобного метода и у знаменитого Иоганна Кеплера (1571—1630) — выдающегося астронома и математика, открытия которого сыграли столь большую роль в создании научной системы мира. Для своего времени Кеплер был, бесспорно, крупнейшим мастером интегрирования (которое тогда и интегрированием-то еще не называлось). В 1614 г. он (вторично) женился; в связи со



ГАЛИЛЕО ГАЛИЛЕЙ



РЕНЕ ДЕКАРТ



ИОГАНН КЕПЛЕР

свадебными торжествами ему пришлось закупить немало вина — и при этом Кеплер убедился в трудности определения (по радиусам оснований бочки и ее высоте) объема винных бочек, имеющих достаточно разнообразную форму. Заинтересовавшись этим вопросом, Кеплер в следующем, 1615 г. выпускает сочинение «Стереометрия винных бочек, преимущественно австрийских, как имеющих самую выгодную форму» («с присоединением дополнения к архимедовой стереометрии», — уведомлял еще читателя автор); здесь он собрал множество задач на определение объемов тел, ограниченных кривыми поверхностями, — формулы для объемов таких тел, ныне получаемые с помощью интегрального исчисления (см. § 7.10), Кеплер находит с большим искусством.

Истинным «предтечей» высшей математики явился француз Рене Декарт (1596—1650), воин и дипломат, естествоиспытатель и абстрактный мыслитель, создатель так называемой *аналитической геометрии* (метода координат в геометрии; ср. гл. 1) и глубокий философ, провозгласивший познаваемость мира и декларировавший основные принципы диалектики, роль в жизни всевозможных процессов, изучение которых, по Декарту, составляет основную задачу математики. При этом Декарт глубоко усовершенствовал математическую символику и математический язык, придав ему современный вид; это сыграло очень большую роль в дальнейшем прогрессе и демократизации математических знаний.

Современником и частично соперником Декарта был еще один французский ученый — юрист по профессии и глубокий математик-любитель Пьер Ферма (1601—1665). Ферма независимо от Декарта и даже несколько раньше его (хоть опубликовано это было позже) также разработал систему использования метода координат и алгебраических выкладок в геометрии, которая ныне называется аналитической геометрией. При этом трактовка темы у Ферма была, пожалуй, более близкой к современной, чем у Декарта, — и если значение сочинений Декарта в истории науки оказалось гораздо бóльшим, то это связано в первую очередь с педагогической активностью Декарта и с более совершенной системой его записей (с подобным же положением мы столкнемся и говоря о создании дифференциального и интегрального исчислений). Одновременно



ПЬЕР ФЕРМА

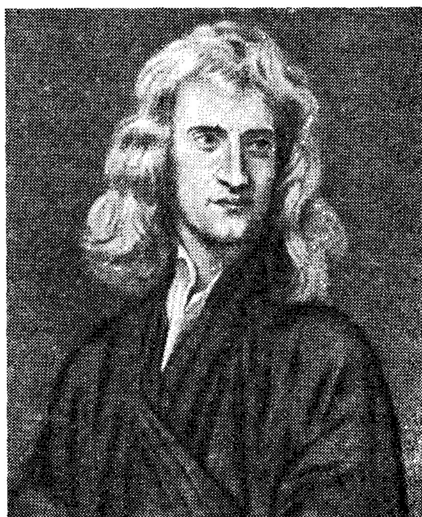


ХРИСТИАН ГЮЙГЕНС

с этим Ферма занимался и теми вопросами, которые ныне относят к области математического анализа: ему принадлежат понятие *дифференциала* (см. ниже § 4.1) и общее замечание о том, что *точкам максимума и минимума (гладкой) функции должно отвечать равенство нулю ее производной* (это утверждение сегодня называют теоремой Ферма), а также остроумные вычисления величин некоторых интегралов (например, намеченный в упр. 3.2.3 метод вывода формулы (5.2.1)). Отдельные результаты в области анализа имел также и Декарт.

Младшим современником Декарта и Ферма был голландец Христиан Гюйгенс (1629—1695), создатель волновой теории света, прославившийся своими работами по применению математики к механике и физике (ему принадлежит, например, строгая математическая теория маятниковых часов; (см. § 7.9). Однако в своих исследованиях Гюйгенс пользовался архаическими приемами древних, в частности Архимеда, считая, что более новая методика не дает никаких преимуществ, ибо любую задачу он может решить «по старинке» (но для того чтобы таким способом свободно решать многочисленные трудные задачи, надо было быть Гюйгенсом).

Истинными создателями высшей математики по праву считаются крупнейшие мыслители XVII в. англичанин Исаак Ньютон (1643—1727) и немец Готфрид Вильгельм Лейбниц (1646—1716); оба они безусловно принадлежат к числу самых глубоких ученых, которых знает вся история науки. Именно им принадлежит связное изложение



ИСААК НЬЮТОН

нового исчисления, цепочка формул, позволяющих без труда найти производную любой алгебраически заданной функции, и полное понимание связи между производной и интегралом и значения этой связи, доставляющей общий алгоритм вычисления интегралов путем обращения к списку производных (см. гл. 4 и 5).

Г. В. Лейбниц был поистине универсальным ученым: его научные интересы включали философию, филологию, историю, психологию (где он был одним из пионеров проникновения в сферы «бессознательного»); биологию (он является одним из предшественников эволюционного учения), геологию и горное



ГОТФРИД ВИЛЬГЕЛЬМ ЛЕЙБНИЦ

дело; математику и механику (в механике Лейбницу принадлежит понятие «живой силы», т. е. кинетической энергии); наряду с этим он был весьма активен как политик и дипломат (он стремился к примирению немецких княжеств, видя в перспективе их объединение в единое государство, и мечтал об объединении католической церкви с протестантской) и как организатор научных академий; в частности, он явился создателем и первым президентом Прусской академии наук в Берлине и несколько раз беседовал с русским царем Петром I о создании Российской академии наук (которая впоследствии была учреждена в полном соответствии с его предложениями и пожеланиями). Математический анализ имел у Лейбница, которому, в частности, принадлежат и современные термины «производная» (Ableitung, по-немецки) и «интеграл»[5], форму, достаточно близкую к принятой в настоящей книге. Производную функции $y=f(x)$ Лейбниц обозначал символом $\frac{dy}{dx}$; при этом «дифференциалы» dy и dx он понимал как «предельные» значения приращений Δy и Δx, к которым мы приходим при неограниченном уменьшении Δx (слова «предел» и соответствующего понятия у Лейбница еще не было). Этот подход отвечал введению производной как тангенса угла, образованного касательной к графику функции с осью Ox: малому Δx отвечает малый треугольник ABC (см. рис. 7.8.1), где $\operatorname{tg} \alpha_1 = \frac{BC}{AB} = \frac{\Delta y}{\Delta x}$; когда Δx неограниченно уменьшается, приращения Δx и Δy заменяются дифференциалами dx и dy, а секущая AC длины ds («дифференциал длины дуги»; см. § 7.8) обращается в касательную (такой треугольник Лейбниц называл *характеристическим*).

Лейбниц, всячески подчеркивающий алгоритмическую сторону нового исчисления, систему правил, автоматически гарантирующих правильный результат при нахождении производных, разработал также методику обращения с дифференциалами, о которой мы еще скажем в гл. 4; использование этой методики и доставляет рецептуру вычисления производных. Интеграл от функции $y=f(x)$ Лейбниц обозначал современным символом $\int f(x)\,dx$.

Неустанная педагогическая деятельность Лейбница и удачная система обозначений и терминов привели к тому, что человечество восприняло высшую математику в созданном им виде. Наряду с этим Лейбниц считается классиком философии (здесь он во многом продолжил и углубил Декарта); ему принадлежит также ряд глубоких идей, частично реализованных лишь в XIX и XX вв., например идея математических машин; мысль о «геометрическом исчислении», из которой впоследствии родилось современное векторное исчисление; попытки «алгоритмизации мышления», при которой, как писал Лейбниц, спорящие более не будут нуждаться в длительных прениях, поскольку один из них всегда сможет сказать другому: «Ну что ж, проверим, кто из нас прав, — вычислим, милостивый государь» (наброски такого «исчисления высказываний», оставшиеся в бумагах Лейбница, очень близки к математической логике XIX и XX вв.).

Совершенно независимо от Лейбница и даже несколько раньше к тем же идеям пришел и И. Ньютон — и одновременность этого великого открытия, независимо сделанного двумя учеными, во многом различными между собой (что сказалось и на путях, какими пришли они к дифференциальному и интегральному исчислениям[6]), лучше всего демонстри-

[5] Первоначально Лейбниц говорил просто *сумма* (и «сумматорное исчисление» вместо «интегрального»); термины *интеграл* (от латинского integer — целый) и *интегральное исчисление* были предложены его учениками, братьями Якобом и Иоганном Бернулли, и с энтузиазмом приняты Лейбницем.

[6] Философ Лейбниц во многом руководствовался тем, что он называл «метафизикой бесконечно малых» (дифференциалов); его вели созданный им язык, развитая символика и терминология, свойственные новому исчислению. Физик же Ньютон исходил из содер-

рует своевременность и неизбежность грандиозной научной революции XVII в.! Ньютон был, пожалуй, больше физиком и астрономом, чем математиком; его роль в создании физики и в становлении научного метода в естествознании никак не может быть преувеличена. Крупнейший математик и механик XVIII—XIX вв. Жозеф Луи Лагранж (1736—1813), с которым мы еще встретимся ниже, сказал как-то о Ньютоне: «Вот счастливейший из всех людей: ведь система мира может быть открыта только единожды». Тот же смысл имеет и знаменитое двустишие современника Ньютона Александра Попа (1688 — 1744), известное у нас в переводе Маршака:

«Был этот мир глубокой тьмой окутан,
Да будет свет! И вот явился Ньютон.»

Для Ньютона производная, конечно, отождествлялась со скоростью: свойства производной воспринимались им как физические свойства скоростей. При этом Ньютону не чужда была и формально-математическая теория производных (и интегралов), основанная на некотором варианте теории пределов. Однако при отсутствии самого определения предела она не делала его построения более безукоризненными, чем (логически не бесспорные) манипуляции Лейбница с бесконечно малыми, а лишь утяжеляла их; возможно, что с этим связано и то обстоятельство, что влияние Лейбница на европейскую математику явно превзошло роль Ньютона.

Производную Ньютон называл *флюксией*, а исходную функцию, для которой вычисляется производная, — *флюентой* (от латинского слова *fluere* — течь; этим подчеркивалось, что рассматриваемые величины являются переменными); при этом флюксия возникала как скорость изменения флюенты, а флюента восстанавливалась по флюксии как путь по скорости. И свое изложение анализа Ньютон начинает с двух основных задач, к которым сводятся все остальные.

1. Длина пройденного пути дана, требуется найти скорость движения в предложенное время.

2. Скорость движения дана, требуется найти длину пройденного в предложенное время пути.

Это, очевидно, задачи нахождения производной заданной функции и нахождения функции по ее производной, т. е. задача вычисления неопределенного интеграла. То, что эти две задачи (в геометрической их постановке — проведение касательной к данной кривой и вычисление ограниченной этой кривой площади) являются взаимно-обратными, было, видимо, впервые обнаружено учителем Ньютона Исааком Барроу (1630—1677), который впоследствии (редкий случай в истории науки!) отказался от должности профессора Кембриджского университета в пользу своего гениального ученика, считая, что Ньютон более его достоин быть профессором. Однако сегодня предложение о связи между производными и интегралами не случайно называют не именем Барроу, а именами Ньютона и Лейбница — ведь только эти два великих ученых поняли всю силу обнаруженного Барроу (и открытого Лейбницем независимо) факта взаимной обратности операций дифференцирования и интегрирования, возможность положить этот факт в основу широкого исчисления производных и интегралов. И именно прозорливости Ньютона и Лейбница математический анализ обязан тем бурным расцветом, который начался сразу за первыми публикациями и выступлениями этих двух великих ученых и который поистине означал наступление новой эры — эры великих научных открытий, бурного штурма тайн природы на благо человека. При этом в отношении попыток приложения дифференциального и интегрального исчисления к раскрытию законов природы приоритет, бесспорно, принадлежит великому естествоиспытателю Ньютону, от которого идет общая идея о том, что законы природы должны иметь форму *дифференциальных уравнений*, связывающих те функции, которые описывают рассматриваемое явление. Ньютону принадлежит формулировка основных уравнений движения (об этом мы подробнее скажем в гл. 9 и 10), которые он затем применил к выводу закона тяготения, к изучению движений небесных тел и к довольно сложным задачам небесной механики (ср. с процитированным выше высказыванием Лагранжа о Ньютоне). Именно с работ Ньютона, по существу, родилась основанная на использовании глубокого математического аппарата теоретическая физика, ныне в своей содержательной части далеко вышедшая за пределы решавшихся Ньютоном задач, а в математической части опирающаяся на весь массив современной науки

жательного смысла новых понятий и их роли в естествознании. (Заметим, что аналитическую геометрию тоже создали одновременно и независимо Декарт и Ферма, причем эти ученые также принадлежали к разным психологическим типам: Декарта здесь можно сравнить с Лейбницем, а весьма глубоко чувствующего физику Ферма (ему, например, принадлежит замечательный принцип, утверждающий, что *траектория света всегда такова, что она обеспечивает быстрейшее его распространение*) — с Ньютоном.)

многократно превышающий созданное Ньютоном и Лейбницем дифференциальное и интегральное исчисление.

Обозначал флюксию (производную) Ньютон точкой, поставленной над буквой, символизирующей исходную функцию: так, производную функции y он записывал как $\dot{y}$. (Сегодня это обозначение производной сохранилось в механике; мы, однако, им пользоваться не будем.) Операцию перехода от флюксии к флюенте Ньютон предложил обозначать точкой, поставленной снизу: так, если $y=x^2$ и $y_1=2x$, то в этой системе обозначений $\dot{y}=y_1$ и $y_1=y$. Такие обозначения подкупают симметрией; теорема о связи между производной и интегралом здесь может быть записана так: $\dot{y}=y$ (где выражение $\dot{y}$ можно понимать двояко — как флюксию от функции y и как флюенту от функции $\dot{y}$). Однако ныне эта красивая запись имеет лишь чисто историческое значение (да и сам Ньютон не был здесь особенно последователен). Обозначение y' для производной функции было введено французским математиком Огюстеном Коши, с именем которого мы еще встретимся ниже.

Глава 4

ВЫЧИСЛЕНИЕ ПРОИЗВОДНЫХ

Удобные и наглядные способы записи тех или иных соотношений и простые правила (алгоритмы), позволяющие механически, без размышлений, производить безошибочные вычисления, играют очень большую роль и в процессе обучения математике, и в развитии самой математики. Так, выдающееся место в истории математики Декарта или Лейбница было связано не только с их математическими достижениями, но и с разработанными этими учеными новыми удобными способами записи, в частности с созданием Лейбницем языка (исчисления) дифференциалов.

В гл. 2 мы разобрали смысл понятия производной и привели первые простейшие примеры нахождения производных. Задачей настоящей главы является разработка общих правил вычисления производных от различных функций: от многочленов; рациональных функций (функций, представимых в виде отношения многочленов); алгебраических функций, содержащих неизвестное под знаком корня (содержащих дробные степени неизвестного); показательных и логарифмических функций; тригонометрических функций и др. Мы укажем, как, зная производные тех или иных функций, найти производные определенных их комбинаций: суммы функций, произведения или частного функций, сложной функции. Таблица производных в конце книги (см. Приложение I) подытожит содержание § 1—12.

§ 1. Дифференциал

Из определения производной следует такой способ (алгоритм) ее вычисления: нужно задаться каким-то значением x и приращением Δx, найти $f(x)$ и $f(x+\Delta x)$, найти приращение $\Delta f=$ $=f(x+\Delta x)-f(x)$, составить отношение $\frac{\Delta f}{\Delta x}$ и перейти к пределу при $\Delta x\to 0$. Однако формула для отношения $\frac{\Delta f}{\Delta x}$, где Δx — произвольное фиксированное число, как правило, оказывается сложнее формулы для предела $\lim\limits_{\Delta x\to 0}\frac{\Delta f}{\Delta x}=\frac{df}{dx}$, т. е. для производной.

Поэтому в дальнейшем мы часто будем писать такие формулы, которые справедливы только в пределе, при приращении, стремящемся к нулю. При малых Δx эти формулы будут «почти справедливы», т. е. точные равенства в них заменятся приближенными, а при больших Δx они могут оказаться и вовсе неверными. Чтобы это подчеркнуть, вместо Δy, Δx мы будем писать dy, dx. Нам нужно только выработать такие правила действий с величинами dy, dx, называемыми *дифференциалами*[1], которые обеспечат выполнение основного равенства

$$\frac{dy}{dx}=y'(x). \qquad (4.1.1)$$

Заметим, что равенство (1) фигурировало у нас уже в гл. 2. Там, однако, оно не было истинным равенством, а лишь условием об обозначениях — оно означало, что записи $y'(x)$ и dy/dx имеют один и тот же смысл. Теперь же мы хотим рассматривать (1) как настоящее (содержательное) равенство, указывающее связь между производной $y'(x)$ функции и дифференциалами dy и dx самой функции y и независимого переменного x: производная $y'(x)$ должна быть равна отношению dy/dx дифференциалов.

Выше было получено приближенное выражение для приращения

$$y(x+\Delta x)-y(x)=\Delta y\approx y'\Delta x. \qquad (4.1.2)$$

Можно сказать, что (2) «становится точным в пределе», при $\Delta x\to 0$. Здесь слова «точно в пределе» не сводятся, разумеется, лишь к тому, что при $\Delta x=$ $=0$ левая и правая стороны (2) совпадают (равны нулю) — они подчерки-

[1] От латинского слова differentia — разность, приращение; таким образом, русским словом «приращение» мы называем величины Δy и Δx, а латинским термином «дифференциал» — «почти приращения» dy и dx.

вают, что при очень малых Δx левая и правая части (2) «почти равны» в том смысле, что их разность гораздо меньше самих этих выражений. В самом деле, разность левой и правой частей (2) при малых Δx имеет порядок малости $(\Delta x)^2$ или еще выше (см. конец § 2.4), так что она действительно (разумеется, при $y'\neq 0$) гораздо меньше величин Δy и $y'\,\Delta x$, имеющих первый порядок малости: даже разделив обе части (2) на очень малое число Δx, мы все равно придем к «почти точному» равенству.

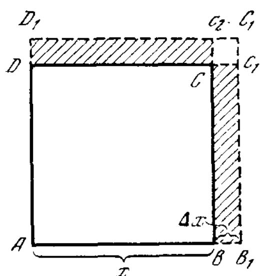


Рис. 4.1.1

Условимся теперь заменять (2) т о ч н ы м равенством

$$dy = y'(x)\cdot dx. \tag{4.1.2а}$$

Поясним это подробнее. Мы будем считать, что *дифференциал независимой переменной* dx, по определению, не отличается от приращения Δx. После этого *дифференциал* dy *функции* $y=y(x)$ определим (точным!) равенством (2а) [2], которое можно дополнить так:

$$dy = y'(x)\cdot dx = y'(x)\cdot \Delta x. \tag{4.1.2б}$$

Таким образом, дифференциал dy функции $y=y(x)$ отличается от ее приращения Δy — ведь для приращения Δy родственное (2б) равенство (2) выполняется лишь приближенно: $\Delta y\approx y'(x)\Delta x$, в то время как точное равенство для приращения $\Delta y=$ $=f(x+\Delta x)-f(x)$ имеет вид

$$\Delta y = y'(x)\,\Delta x+\sigma, \tag{4.1.3}$$

где σ — величина более высокого порядка малости, чем Δx. Именно это мы и имели в виду, говоря, что равенство (2), или

$$\frac{\Delta y}{\Delta x}\approx y'(x),$$

является «точным в пределе»: само по себе отношение $\frac{\Delta y}{\Delta x}$, вообще говоря, отлично от $y'(x)$, но вот $\lim\limits_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=y'(x)$, поскольку при малых Δx слагаемым σ в правой части (3) можно пренебречь. Однако эти слова о пределе, о пренебрежении какой-то частью формулы становятся излишними при переходе от «языка приращений» к «языку дифференциалов»: соотношения (1) и (2а), (2б), использующие дифференциалы, являются уже обычными (точными) равенствами.

Все это подробно обсуждалось в § 2.4 в связи с вопросом о приближенном вычислении значений функций (с. 65 и след.). При вычислении, скажем, значений функции $y=x^2$ или $y_1=x^3$ мы с полным основанием полагаем, что если Δx мало, то $\Delta y\approx 2x\,\Delta x$, соответственно $\Delta y_1\approx 3x^2\,\Delta x$, поскольку сделанная при этом ошибка будет неощутимо мала. Так, приближенное равенство $\Delta y\approx 2x\,\Delta x$ означает, что приращение Δy площади ($y=x^2$) квадрата $ABCD$ со стороной $AB=x$ (рис. 1) при изменении длины его стороны на Δx можно заменить суммой площадей двух прямоугольников BB_1c_1C и DD_1c_2C, пренебрегая при малом $BB_1=\Delta x$ (гораздо меньшим их!) квадратиком $Cc_1C_1c_2$ площади $(\Delta x)^2$. Обращаясь к функции $y_1=x^3$ и полагая, скажем, $x=2$, $y_1(x)=8$, составим следующую таблицу приращений $\Delta y_1=y_1(x+\Delta x)-y_1(x)$ и дифференциалов $dy_1=y_1'(x)\,\Delta x$ этой функции, а также погрешностей $\frac{\sigma}{\Delta y_1}=\frac{\Delta y_1-dy_1}{\Delta y_1}$ приближенной формулы (2):

Δx	1	0,5	0,2	0,1
Δy_1	19	7,625	2,648	1,261
dy_1	12	6	2,4	1,2
$\sigma/\Delta y_1$	37%	21%	9%	5%

Δx	0,05	0,01	0,001
Δy_1	0,615	0,1206	0,012006
dy_1	0,6	0,12	0,012
$\sigma/\Delta y_1$	2,5%	0,5%	0,05%

[2] Заметьте, что в то время как приближенное равенство (2) является т е о р е м о й, которая нуждается в доказательстве (нуждается в доказательстве тот факт, что величина σ в правой части (3) имеет более высокий порядок малости, чем Δx или Δy), равенство (2а) (и (2б)) является о п р е д е л е н и е м — оно определяет величину dy.

Эта таблица весьма наглядно иллюстрирует утверждение о том, что формула (2) «точна в пределе».

Более полные сведения об «остаточном члене» σ приближенной формулы (2) будут получены в гл. 6. Там мы докажем, что σ можно представить в виде суммы $a(\Delta x)^2+$ $+b(\Delta x)^3+\ldots$, коэффициенты a, b, ... которой не зависят от Δx. Но при малых приращениях Δx все степени $(\Delta x)^2$, $(\Delta x)^3$ и т. д. гораздо меньше самой величины Δx; поэтому $|\sigma|=$ $=|a(\Delta x)^2+b(\Delta x)^3+\ldots|$ гораздо меньше (пропорционального $|\Delta x|$!) дифференциала $|dy|=|y'(x)\Delta x|$. Разумеется, последнее утверждение теряет силу при $y'(x)=0$, когда дифференциал dy обращается в нуль и меньше его положительная величина $|\sigma|$ быть никак не может; однако обычно и здесь можно пользоваться приближением (2), пренебрегая крайне малой ошибкой σ. Случаи же отсутствия у функции $y=y(x)$ производной (и дифференциала) (ср., например, § 7.2) имеют «экзотический характер», — и мы рекомендуем читателю пока просто игнорировать их.

Понятие дифференциала и запись $\frac{dy}{dx}$ для производной были введены Лейбницем, который при этом не слишком задумывался над содержанием понятия «дифференциал»; гораздо более интересовали его правила оперирования с дифференциалами, составившие лейбницевское «исчисление дифференциалов». Нетрудно, однако, придать понятию дифференциала точный смысл и даже проиллюстрировать его на чертеже.

Равенство (3) можно переписать так:

$$\Delta y = dy + \sigma; \tag{4.1.3а}$$

с ним связана распространенная формулировка: *дифференциал функции — это главная линейная часть ее приращения*. Другими словами, дифференциал $dy=y'(x)\Delta x$ $(=y'(x)\,dx)$ является л и н е й н о й функцией приращения Δx, в то время как приращение Δy, вообще говоря, зависит от Δx довольно сложно (не линейно) [3]. С другой стороны, dy представляет собой «главную часть» приращения Δy функции в том смысле, что при малых Δx второе слагаемое в правой части (3а) гораздо меньше первого.

Вспомним теперь, что производная $y'(x)$ равна тангенсу угла α наклона касательной к графику функции $y=$ $=y(x)$ (см. § 2.5). Поэтому, если на рис. 2 $PQ=\Delta x$, то

$$NQ_1 = \operatorname{tg}\alpha\cdot\Delta x = y'(x)\,\Delta x = dy,$$

в то время как приращение функции $\Delta y=y(x+\Delta x)-y(x)$ изображается отрезком M_1Q_1. Из рис. 2 видно, что при малых Δx дифференциал dy будет близок к приращению Δy функции,

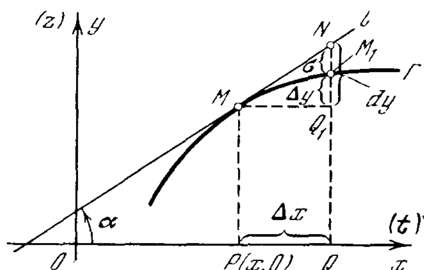


Рис. 4.1.2

в то время как при больших Δx это уже, вообще говоря, будет не так. Однако такое графическое представление дифференциала функции нужно лишь для правильного понимания сути использования дифференциала в приближенных вычислениях, в несколько иной форме (без использования самого слова «дифференциал») описанного в § 2.4. В этой же главе нам нужна будет лишь техника работы с дифференциалами, символическое «исчисление дифференциалов»; точный же смысл понятия «дифференциал», связанный с рис. 2, использоваться нигде не будет, — и читатель вполне может просто о нем забыть.

Понятию дифференциала можно также придать прозрачный механический смысл. Предположим, что на рис. 2 абсцисса x — это время (именно этот смысл имеет буква t в скобках у конца оси абсцисс), а ордината y (или z) — путь; нас интересует процесс изменения пути со временем $y=y(x)$, или, в привычных нам по гл. 2 и 3 обозначениях, $z=z(t)$. Представление о постоянно меняющейся под влиянием каких-то сил скорости не слишком просто; поэтому при изучении движения

[3] Дифференциал $dy=y'(x)\Delta x$ функции зависит от д в у х чисел (переменных) x и Δx — он является ф у н к ц и е й д в у х п е р е м е н н ы х (линейной по переменной Δx, но довольно сложно, вообще говоря, зависящей от переменной x).

в окрестности какого-то момента времени (положение тела в этот момент изображается точкой M графика движения) удобно считать, что начиная с этого момента скорость перестала изменяться (это предположение равносильно гипотезе о том, что в рассматриваемый момент времени мы «отключили» все действующие на тело силы, предоставив ему далее двигаться по инерции, т. е. с постоянной скоростью, см. гл. 9). Тогда начиная с этого момента x скорость все время будет оставаться равной мгновенной скорости $v(x)=\frac{dy}{dx}$ в момент x $\left(\text{или } \frac{dz}{dt}\text{ в момент } t\right)$, и пройденный за время Δx путь будет равен $v(x)\Delta x=y'\Delta x=dy$. (На рис. 2 равномерному движению тела отвечает прямая l, в то время как графиком исходного — неравномерного — движения служит линия Г.) При малых Δx (или Δt) этот гипотетический путь NQ_1 $(=dy$ или $dz)$ будет отличаться от истинного пути M_1Q_1 $(=\Delta y$ или $\Delta z)$ весьма мало — на малую величину $NM_1=\sigma$ более высокого порядка, чем PQ $(=\Delta x$ или $\Delta t)$ [4].

Правила пользования дифференциалами должны быть такими, чтобы выполнялось соотношение (1): для этого в формулах для дифференциалов нужно сразу отбрасывать члены, пропорциональные $(dx)^2$ или более высоким степеням dx.

Рассмотрим один простейший пример; сравним на нем технику работы с приращениями и с дифференциалами. Пусть мы имеем функцию $y=x^2$. Раньше мы писали

$$\Delta y=(x+\Delta x)^2-x^2=2x\Delta x+(\Delta x)^2, \quad (4.1.4)$$

откуда

$$\frac{\Delta y}{\Delta x}=2x+\Delta x \quad \text{и} \quad y'=\lim_{\Delta x\to 0}\frac{\Delta y}{\Delta x}=2x.$$

С помощью дифференциалов запишем

$$dy=(x+dx)^2-x^2=2xdx, \quad (4.1.4а)$$

т. е. член $(dx)^2$ в правой части (4) мы сразу отбросили!

Смысл такого оперирования с дифференциалами теперь понятен. Разумеется, $\Delta y=(x+dx)^2-x^2=2xdx+(dx)^2$ (см. (4); напоминаем, что dx и Δx — это одно и то же). Когда же мы переходим от Δy к dy, мы оставляем лишь «главную линейную часть» Δy, т. е. отбрасываем члены, пропорциональные Δx (или dx) во второй и высших степенях, сохраняя лишь члены, пропорциональные первой степени dx. Эти правила «исчисления дифференциалов», если только привыкнуть к ним, очень упрощают все выкладки; в последующих параграфах мы будем широко ими пользоваться.

Остановимся, наконец, на одном замечательном свойстве дифференциалов, которое знал (и высоко ценил) еще Лейбниц. Рассмотрим какую-либо функцию, скажем $y=x^2$; при этом

$$y'=2x \quad \text{и} \quad dy=2x\Delta x. \quad (4.1.5)$$

Пусть теперь x — не независимая переменная, а вспомогательное неизвестное, зависящее от независимой переменной t (можно считать, что t является временем). Сохранят ли при этом силу соотношения (5)? Нетрудно понять, что не сохранят. Так, если $x=t^2$, то производная функции $y=x^2=t^4$ равна $4t^3$, что отлично от $2x=2t^2$. Во второй же из формул (5) Δx теперь равно $(t+\Delta t)^2-t^2=2t\Delta t+(\Delta t)^2$, поэтому

$$2x\Delta x=2t^2[2t\Delta t+(\Delta t)^2]=4t^3\Delta t+2t^2(\Delta t)^2,$$

что отлично от выражения $dy=4t^3\Delta t$, стоящего в левой части той же формулы. Но вот основное соотношение (2а) сохраняет силу и в новых условиях: ведь здесь $dx=2t\Delta t$, так что $2xdx=2t^2(2t\Delta t)=4t^3\Delta t=dy$.

Является ли это обстоятельство случайным? Нет, не является: формула (2а) сохраняет силу и в том случае, когда x есть независимая переменная, и в ситуации, когда x есть («промежуточная») функция независимой переменной t (ср. упр. 1 и 2).

---

[4] Именно в таком (механическом) обличьи появился дифференциал у Ньютона, который назвал его термином «момент»; при этом Ньютон пришел к этому понятию раньше Лейбница, который заговорил о дифференциалах лишь с 1675 г., в то время как «моменты функций» (или «флюент», как предпочитал говорить Ньютон) появились у Ньютона уже в 1666 г. Еще раньше, начиная примерно с конца 30-х годов XVII в., пользовался дифференциалами (без специального наименования и обозначения этих величин) Ферма; именно на этих понятиях была основана его известная теорема о точках максимума и минимума. Но развернутое «исчисление дифференциалов», безусловно, является заслугой Лейбница.

В силу этого работать с дифференциалами зачастую проще и удобнее, чем с производными и с приращениями, где всегда необходимо помнить о характере аргумента функции (независимая переменная? вспомогательная функция?). Многочисленные примеры этого мы будем иметь в дальнейшем.

Обосновать сформулированное свойство дифференциалов проще всего с помощью формулы (3), где α есть величина более высокого порядка малости, чем dx (и $dy=A\,dx$). В силу этой формулы

$$\Delta y = B\,dt + \alpha, \qquad (4.1.6)$$

где $B\,dt = dy$, а α малò. С другой стороны,

$$\Delta x = C\,dt + \beta, \qquad (4.1.6а)$$

где $C\,dt = dx$, а β малò, и если производная $\frac{dy}{dx}$ в рассматриваемой точке (при $t = t_0$ или при $x = x(t_0) = x_0$) равна D, то

$$\Delta y = D\Delta x + \gamma, \qquad (4.1.6б)$$

где γ малò. Но из (6а), (6б) следует, что $D\,dx$ $(=DC\,\Delta t)$ совпадает с дифференциалом функции $y=y(t)$: ведь это, во-первых, линейная функция Δt, а, во-вторых, разность

$$\Delta y - D\,dx = (D\Delta x + \gamma) - D\,dx =$$
$$= [D(dx + \beta) + \gamma] - D\,dx = D\beta + \gamma$$

является малой величиной (величиной более высокого порядка малости, чем Δt), поскольку таковыми являются величины β и γ, а $D\,dx$ — это, конечно, $\frac{dy}{dx}\,dx$.

Вторую производную $y''(x)$ функции $y=y(x)$ мы обозначали выше (см. § 2.7) через

$$y''(x) = \frac{d^2y}{dx^2}. \qquad (4.1.7)$$

Эту запись можно оправдать с помощью соображений, родственных тем, которые мы развивали в связи с записью $y' = \frac{dy}{dx}$.

Назовем величину $y''(x)(\Delta x)^2$ $(=y''(x)(dx)^2)$ *вторым дифференциалом* функции $y(x)$ и обозначим ее через d^2y; в таком случае обозначение (7) для второй производной $y''(x)$ можно будет трактовать как отношение $d^2y : (dx)^2$ второго дифференциала d^2y функции y к квадрату $(dx)^2$ (первого) дифференциала (приращения) аргумента x. Второй дифференциал d^2y функции $y=y(x)$ можно также изобразить на чертеже, что открывает пути к использованию его для приближенного вычисления значений функции. Условимся временно писать x_0 вместо x и x вместо $x+\Delta x$ и найдем параболу П:

$$Y = a(x - x_0)^2 + b(x - x_0) + c, \qquad (4.1.8)$$

наиболее тесно примыкающую к графику Г функции $y=y(x)$, т. е. такую, чтобы разность $Y(x) - y(x)$ при малых $x - x_0 = \Delta x$ была возможно меньше [5]. Задача нахождения такой параболы (задача о «наилучшем квадратичном приближении» (8) функции $y(x)$) совпадает с рассматриваемой в гл. 6; из результатов этой главы следует, что коэффициенты уравнения (8) параболы П таковы: $a = {}^1/_2\, y''(x_0)$, $b = y'(x_0)$, $c = y(x_0)$. В этом случае разность $Y(x) - y(x)$ при малых Δx будет являться величиной третьего (или еще высшего) порядка малости по отношению к $\Delta x = (x - x_0)$. Если изобразить на одном чертеже (рис. 3) кривую $y=y(x)$, параболу $Y=Y(x)$ и касательную l

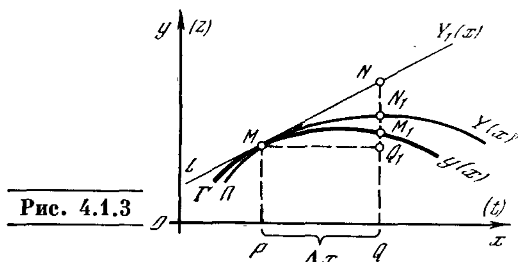


Рис. 4.1.3

$$Y_1 = kx + s \qquad (4.1.9)$$

к линии Г (где $k = y'(x_0)$ и $s = y(x_0) - kx_0$), то прямая $x=\text{const}$ пересечет линии Г, l и П в таких точках M_1, N и N_1, что (см. рис. 3):

$$Q_1M_1 = y(x) - y(x_0) = \Delta y,$$
$$Q_1N = y'(x_0)\,\Delta x = dy,$$
$$Q_1N_1 = {}^1/_2\, y''(x_0)(\Delta x)^2 + y'(x_0)(\Delta x) =$$
$$= {}^1/_2\, d^2y + dy,$$

т. е.

$$Q_1N = dy \quad \text{и} \quad NN_1 = {}^1/_2\, d^2y$$

(в нашем случае d^2y отрицательно).

Рис. 3 убедительно иллюстрирует преимущество, которое имеет приближение $Y = y_0 + dy + {}^1/_2\, d^2y$ $(= y(x_0) + y'(x_0)\,\Delta x + {}^1/_2\, y''(x_0)\,\Delta x^2)$ значения $y(x)$ функции по сравнению с приближением $Y_1 = y_0 + dy$ $(= y(x_0) + y'(x_0)\,\Delta x)$: во втором случае мы заменяем точку M_1

[5] Эту параболу (*соприкасающуюся параболу* П кривой Г; ср. § 7.9) можно описать так: проведем параболу $П_1$ через три близкие точки (x_0, y_0), (x_1, y_1) и (x_2, y_2) кривой Г, т. е. подберем a, b и c в (8) так, чтобы было $Y(x_0) = y_0$, $Y(x_1) = y_1$ и $Y(x_2) = y_2$; тогда при стремлении x_1 и x_2 к x_0 найденная парабола будет стремиться к П (так что П — это предел проведенной параболы $П_2$ при $x_1, x_2 \to x_0$).

на рис. 3 точкой N, а в первом случае—более близкой к M_1 точкой N_1.

Второму дифференциалу d^2y функции y можно, так же как и первому дифференциалу dy, придать и механическое значение. Снова условимся считать, что ось абсцисс на рис. 3 — это ось времени t; по оси ординат (ось z) отложен путь тела. Пусть теперь тело начиная с некоторого момента времени (изображаемого точкой M графика Г) движется равноускоренно — это предположение равносильно гипотезе о том, что начиная с рассматриваемого момента действующие на точку силы перестали изменяться, ибо

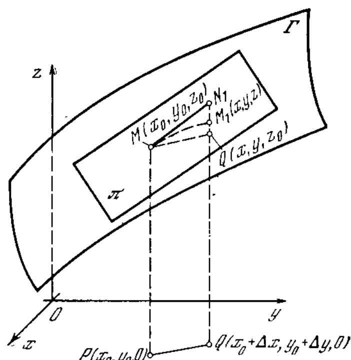


Рис. 4.1.4

именно под действием постоянной силы материальная точка движется равноускоренно (см. гл. 9). В таком случае графиком движения точки станет парабола П, и (гипотетический) путь, пройденный за время Δx (или Δt), будет равен $dy+{}^1/_2d^2y$ (или $dz+{}^1/_2\, d^2z$), на рис. 3 он изобразится отрезком Q_1N_1.

Заметим, наконец, что понятие дифференциала может быть перенесено также и на *функции двух* (или большего числа) *переменных*. Пусть $z=F(x, y)$ — произвольная функция двух переменных x и y; читатель, знакомый с методом координат в пространстве, знает, что ее «графиком» является *поверхность* в пространстве, отнесенном к трем координатам x, y и z (рис. 4).

Рассмотрим некоторые (фиксированные) значения x_0 и y_0 переменных, которым отвечает точка M графика с координатами x_0, y_0 и $z_0=F(x_0, y_0)$. Если теперь координаты x и y близки к x_0 и y_0, т. е. разности $x-x_0=\Delta x$ и $y-y_0=\Delta y$ малы, то значение

$$z=F(x, y)=F(x_0+\Delta x, y_0+\Delta y)$$

можно найти следующим образом.

Будем искать величину $z_1=F(x, y_0)=F(x_0+\Delta x, y_0)$. Поскольку здесь по сравнению с (предполагаемой известной) величиной $z_0=F(x_0, y_0)$ изменился лишь один аргумент x, то в силу сказанного выше

$$z_1=F(x_0+\Delta x, y_0)\approx \approx F(x_0, y_0)+F'_x(x_0, y_0)\,\Delta x, \qquad (4.1.10)$$

где

$$F'_x(x, y)=\frac{\partial F(x, y)}{\partial x}= =\lim_{\Delta x\to 0}\frac{F(x+\Delta x, y)-F(x, y)}{\Delta x}$$

— так называемая *частная производная* функции F по x, получаемая в предположении, что меняется лишь аргумент x функции, а аргумент y закреплен (см. § 12). Далее аналогично получаем, закрепляя на этот раз значение первого аргумента x функции, а меняя один только второй аргумент y:

$$z=F(x_0+\Delta x, y_0+\Delta y)\approx \approx F(x_0+\Delta x_0, y_0)+F'_y(x_0+\Delta x, y_0)\,\Delta y. \qquad (4.1.11)$$

Объединяя приближенные равенства (10) и (11), получаем

$$z=F(x, y)=F(x_0+\Delta x, y_0+\Delta y)\approx \approx F(x_0, y_0)+F'_x(x_0, y_0)\,\Delta x+ +F'_y(x_0+\Delta x, y_0)\,\Delta y.$$

Учитывая близость частной производной $F'_y(x_0+\Delta x, y_0)$ к $F'_y(x_0, y_0)$ и заменяя Δx и Δy на имеющие тот же смысл величины dx и dy (ведь x и y — это независимые переменные, а не функции!), последнее равенство можно переписать еще и так:

$$z=F(x_0+dx, y_0+dy)\approx \approx F(x_0, y_0)+F'_x(x_0, y_0)\,dx+F'_y(x_0, y_0)\,dy= \left(=F(x_0, y_0)+\frac{\partial F(x_0, y_0)}{\partial x}dx+\frac{\partial F(x_0, y_0)}{\partial y}dy\right). \qquad (4.1.12)$$

Фигурирующее в правой части (12) слагаемое $dz=\frac{\partial F}{\partial x}dx+\frac{\partial F}{\partial y}dy$ и называется *дифференциалом функции* $z=F(x, y)$ *двух переменных*. В приближенном равенстве

$$z\approx z_0+dz \qquad (4.1.12а)$$

(здесь $z_0=F(x_0, y_0)$) ошибка представляет собой величину второго порядка малости по сравнению с dx и dy (или Δx и Δy), которые мы здесь полагаем малыми одного порядка. Геометрически равенство (12) (или (12а)) означает, что вместо точки M_1 графика функции F (поверхности Г) мы берем соответствующую M_1 точку N_1 касательной плоскости π к поверхности Г в точке M (см. рис. 4).

Упражнения

4.1.1. Докажите, что $\frac{dy}{dt}=\frac{dy}{dx}\frac{dx}{dt}$, если: а) $y=x^2$, а $x=\sqrt{t}$; б) $y=\sqrt{x}$, а $x=t^2$.

4.1.2. Пусть $y=x^2$, где $x=\sqrt{t+1}$, а $t=u^2$. Докажите, что

$$dy=\frac{dy}{dx}dx=\frac{dy}{dt}dt=\frac{dy}{du}du.$$

§ 2. Производная суммы и произведения функций

В качестве следующего примера использования языка дифференциалов рассмотрим с у м м у двух функций $f(x)$ и $g(x)$, взятых с (произвольными) постоянными коэффициентами A и B:

$$y=Af(x)+Bg(x).$$

Используя дифференциалы, запишем

$$dy\approx y(x+dx)-y(x)=[Af(x+dx)+Bg(x+dx)]-[Af(x)+Bg(x)]=A[f(x+dx)-f(x)]+B[g(x+dx)-g(x)]\approx Adf+Bdg=Af'dx+Bg'dx,$$

откуда

$$y'=\frac{dy}{dx}=Af'+Bg'. \qquad (4.2.1)$$

Читатель легко получит эту формулу, пользуясь не дифференциалами, а приращениями и пределами.

В частности, для суммы и разности двух функций (случаи $A=B=1$ и $A=1$, $B=-1$) имеем:

$$(f+g)'=f'+g'; \quad (f-g)'=f'-g'. \qquad (4.2.2)$$

Найдем теперь производную п р о и з в е д е н и я двух функций $g(x)$ и $h(x)$. Положим

$$f(x)=g(x)h(x).$$

Тогда

$$df(x)\approx f(x+dx)-f(x)=g(x+dx)h(x+dx)-g(x)h(x).$$

Но

$$g(x+dx)\approx g(x)+dg,$$
$$h(x+dx)\approx h(x)+dh.$$

Поэтому

$$df\approx[g(x)+dg][h(x)+dh]-g(x)h(x)=g(x)dh+h(x)dg+dgdh.$$

Заметим, что $dg=g'(x)dx$, $dh=h'(x)dx$, откуда

$$dhdg=g'(x)h'(x)(dx)^2.$$

Величина $dhdg$ пропорциональна $(dx)^2$, поэтому, по правилам работы с дифференциалами, произведением $dhdg$ в выражении для df мы пренебрегаем. Окончательно получаем

$$df=g(x)dh+h(x)dg. \qquad (4.2.3)$$

Разделив все члены (3) на dx, получим

$$\frac{df}{dx}=\frac{d(gh)}{dx}=g\frac{dh}{dx}+h\frac{dg}{dx}. \qquad (4.2.4)$$

Это выражение можно запомнить так: производная произведения gh равна сумме производной, взятой в предположении, что от x зависит только h, а g постоянно $\left(\text{член } g\frac{dh}{dx}\right)$, и производной, взятой в предположении, что от x зависит только g, а h постоянно $\left(\text{член } h\frac{dg}{dx}\right)$. При этом, естественно, само (постоянное) значение g в члене $g\frac{dh}{dx}$ берется при том значении x, при котором ищется значение производной; то же относится и к h во втором слагаемом. Ниже мы увидим, что сходная процедура применима и в некоторых других случаях[1], например в случае функции $y=f(x)^{g(x)}$.

Как мы действовали бы старым способом? Элементарная алгебра дает точное равенство

$$\Delta f=(g+\Delta g)(h+\Delta h)-gh=g\Delta h+h\Delta g+\Delta h\Delta g.$$

Разделим обе части на Δx:

$$\frac{\Delta f}{\Delta x}=g(x)\frac{\Delta h}{\Delta x}+h(x)\frac{\Delta g}{\Delta x}+\frac{\Delta h}{\Delta x}\frac{\Delta g}{\Delta x}\Delta x. \qquad (4.2.5)$$

(Заметим, что последнее слагаемое правой части (5) мы для удобства дополнительно умножили и разделили на Δx.)

До сих пор все равенства являются точными, справедливыми при любых значениях Δx. Перейдем теперь к пределу при $\Delta x\to 0$. При этом будем иметь

$$\lim_{\Delta x\to 0}\frac{\Delta f}{\Delta x}=f', \quad \lim_{\Delta x\to 0}\frac{\Delta h}{\Delta x}=h', \quad \lim_{\Delta x\to 0}\frac{\Delta g}{\Delta x}=g'$$

и в силу (5)

$$f'=gh'+g'h. \qquad (4.2.6)$$

[1] Ясно, что формула для производной суммы функций (если $F=f+g$, то $F'=f'+g'$) также совместима с этим правилом: ведь если считать f постоянным, то производная F' обратится в g', а если постоянно g, то она станет равна f'.

Последний член в формуле (5) при переходе к пределу исчез, так как первые два множителя в пределе дают произведение $h'g'$, а Δx мы устремили к нулю.

С помощью приращений и перехода к пределу получен тот же результат, что и с помощью дифференциалов, но усилий это потребовало больше. Удивляться тут нечему: в случае дифференциалов мы отбросили произведение $dhdg$ механически, на основании общего правила, согласно которому надо отбрасывать члены с $(dx)^2$, $(dx)^3$ и т. д., а значит, и любые произведения двух, трех и большего числа дифференциалов. При расчете с помощью приращений мы, в сущности, по ходу дела доказали это правило еще раз для случая произведения функций.

Последовательные расчеты с помощью приращений нужны для обоснования правил и понимания их. Но после того как понимание достигнуто, пользование дифференциалами быстрее ведет к цели, является более деловым — и было бы смешно каждый раз «танцевать от печки», т. е., решая каждую конкретную задачу, записывать, что производная есть предел отношения, выписывать это отношение и т. д.

Пример. Пусть $f=(3x^2+5)(2x-4)$; требуется найти $f'(x)$, и в частности $f'(0)$.

Здесь

$$g=3x^2+5, \quad \frac{dg}{dx}=3\cdot 2x+0=6x;$$

$$h=2x-4, \quad \frac{dh}{dx}=2-0=2.$$

Поэтому

$$\frac{df}{dx}=(3x^2+5)\cdot 2+(2x-4)\cdot 6x=$$
$$=18x^2-24x+10,$$

и в частности

$$f'(0)=\left.\frac{df}{dx}\right|_{x=0}=10.$$

Правило для нахождения производной произведения легко обобщить и на случай нескольких множителей. Например, для произведения четырех функций $f(x)$, $g(x)$, $h(x)$, $k(x)$ получаем

$$\frac{d(fghk)}{dx}=fgh\frac{dk}{dx}+fgk\frac{dh}{dx}+$$
$$+fhk\frac{dg}{dx}+ghk\frac{df}{dx}. \qquad (4.2.7)$$

Таким образом, здесь мы имеем прежнее правило, согласно которому производную комбинации нескольких функций можно найти как сумму производных, вычисленных в предположении, что каждый раз меняется только одна функция, а остальные сохраняют постоянное значение.

Из полученных нами формул можно сделать дальнейшие выводы. Ясно, что если $f(x)=g(x)+h(x)$ и $f'(x)=g'(x)+h'(x)$, то

$$f''(x)=(f'(x))'=(g'(x)+h'(x))'=$$
$$=g''(x)+h''(x). \qquad (4.2.8)$$

Несколько сложнее выглядит формула для второй производной произведения функций: если $f(x)=g(x)h(x)$, то $f'(x)=g(x)h'(x)+g'(x)h(x)$ и $f''(x)=(g(x)h'(x)+g'(x)h(x))'=(g(x)h''(x)+g'(x)h'(x))+(g'(x)h'(x)+g''(x)h(x))$, т. е.

$$f''(x)=g(x)h''(x)+2g'(x)h'(x)+$$
$$+g''(x)h(x). \qquad (4.2.9)$$

Упражнения

4.2.1. Используйте найденные выше правила для вычисления производных следующих функций: а) $y=x^4$ $(=x^2\cdot x^2)$; б) $y=ax^5+bx^4+cx^3+dx^2+ex+f$; в) $y=\sqrt{x}$, использовав тождество $\sqrt{x}\sqrt{x}=x$. [Указание. Воспользуйтесь равенством $y'y+yy'=x'=1$.]

4.2.2. Выразите третью производную $f'''(x)=(f''(x))'$ функции $f(x)=g(x)h(x)$ через функции f и g и их производные.

4.2.3. Пусть $f(x)=g(x)h(x)k(x)$; чему равна функция $f''(x)$?

§ 3. Сложная функция. Производная частного двух функций

Пусть z задано как функция y, например $z=\frac{1}{y}$, а y в свою очередь есть функция x, например $y=x^2+5$. Очевидно, что каждому значению x соответствует определенное y, а так как каждому y соответствует определенное z, то в конце концов каждому x соответствует определенное z, т. е. z есть функция x. Всегда можно, подставляя выражение y через x, выписать непосредственно формулы для $z(x)$; так, в нашем случае $z=\frac{1}{x^2+5}$.

Но техника дифференцирования (т. е. нахождения производных) облегчается, если свести все функции к соче-

таниям самых простых из них: ведь в нашем случае каждая из зависимостей $z=\frac{1}{y}$ и $y=x^2+5$ в отдельности проще, чем $z=\frac{1}{x^2+5}$. Сводя сложные функции к комбинациям более простых, мы сумеем обойтись правилами нахождения производных этих простых функций.

Найдем дифференциал сложной функции $z\,[y\,(x)]$. Рассматривая z как функцию y, напишем

$$dz=\frac{dz}{dy}dy \tag{4.3.1}$$

(ср. со сказанным в § 1 о записи дифференциала сложной функции). Но y есть функция x; поэтому

$$dy=\frac{dy}{dx}dx. \tag{4.3.2}$$

Подставляя (2) в (1), получим

$$dz=\frac{dz}{dy}\frac{dy}{dx}dx. \tag{4.3.3}$$

Поделив обе части на dx, получим *правило дифференцирования* (определения производной) *сложной функции* [1]:

$$\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}. \tag{4.3.3а}$$

Вид этой формулы вполне соответствует тому, что было сказано о возможности обращения с дифференциалами как с обычными алгебраическими величинами: в произведении $\frac{dz}{dy}\frac{dy}{dx}$ можно сократить dy.

Напомним, что z задано как функция y; поэтому $\frac{dz}{dy}$ также является функцией y. Но так как само y есть функция x, то, подставляя $y=y\,(x)$ в выражение $\frac{dz}{dy}$, мы выразим $\frac{dz}{dy}$, а следовательно, и $\frac{dz}{dx}$ как функцию x.

Проделаем расчет для одного частного случая, который нам понадобится в дальнейшем. Пусть

$$z=\frac{1}{y\,(x)}. \tag{4.3.4}$$

[1] Запись z' вместо $\frac{dz}{dx}$ в этом случае употреблять нельзя, поскольку неясно, что имеется в виду: $\frac{dz}{dx}$ или $\frac{dz}{dy}$. (Можно здесь писать z'_x или z'_y.)

Мы знаем, что если $z=\frac{1}{y}$, то $\frac{dz}{dy}=-\frac{1}{y^2}$. Значит,

$$dz=-\frac{1}{y^2}dy=-\frac{1}{y^2}\frac{dy}{dx}dx$$

и

$$\frac{dz}{dx}=-\frac{1}{y^2}\frac{dy}{dx}. \tag{4.3.5}$$

Так, например, если $y=x^2+5$, а $z=\frac{1}{y}=\frac{1}{x^2+5}$, то

$$\frac{dz}{dx}=-\frac{1}{(x^2+5)^2}\frac{d\,(x^2+5)}{dx}=-\frac{2x}{(x^2+5)^2}.$$

Вот еще один простой пример: пусть $y=x^{3/2}=\sqrt{x^3}$. Эту формулу можно переписать так: $y=\sqrt{u}$, где $u=x^3$. Согласно (3а) имеем

$$\frac{dy}{dx}=\frac{dy}{du}\frac{du}{dx}=\frac{1}{2\sqrt{u}}3x^2=\frac{3x^2}{2\sqrt{x^3}}=\frac{3}{2}\sqrt{x}.$$

Ранее этот же результат был получен более сложным путем (см. упр. 2.4.3).

Правило (3а) вычисления производной сложной функции сохраняется и при «многоступенчатой» зависимости. Так, если $z=z\,(y)$, $y=y\,(x)$, $x=x\,(t)$, а $t=t\,(w)$, то

$$\frac{dz}{dw}=\frac{dz}{dy}\frac{dy}{dx}\frac{dx}{dt}\frac{dt}{dw}. \tag{4.3.6}$$

С помощью (5) легко найти производную ч а с т н о г о (отношения) $f=\frac{h}{g}$ двух функций h и g. Для этого запишем f в виде произведения: $f=h\frac{1}{g}$. Отсюда имеем

$$f'=h\left(\frac{1}{g}\right)'+h'\frac{1}{g}. \tag{4.3.7}$$

Но (см. (5)) $\left(\frac{1}{g}\right)'=-\frac{1}{g^2}g'$. Подставляя это выражение в (7), получим требуемый результат:

$$f'=\left(\frac{h}{g}\right)'=-\frac{h}{g^2}g'+h'\frac{1}{g},$$

или [2]

$$\left(\frac{h}{g}\right)'=\frac{h'g-hg'}{g^2}. \tag{4.3.8}$$

[2] Ясно, что и формула (8) отвечает сформулированному в § 4.2 общему принципу: $\left(\frac{h}{g}\right)'=\frac{h'}{g}+h\left(\frac{1}{g}\right)'$.

Упражнения

4.3.1. Найдите производную от $z=(ax+b)^2$ как производную сложной функции $z=y^2$, где $y=ax+b$. Раскройте в выражении $(ax+b)^2$ скобки и найдите ту же производную непосредственно.

4.3.2. Найдите производные функций: а) $z=\frac{1}{ax+b}$; б) $z=\frac{1}{(ax+b)^2}$; в) $z=\frac{1}{1+1/x}$; г) $y=\frac{x^3+5x^2}{x+1}$; д) $y=\frac{x-1}{x^2+2}$.

4.3.3. Пусть $y=\frac{f(x)}{g(x)}$; выразите вторую производную $\frac{d^2y}{dx^2}$ функции y через функции f и g и их (первые и вторые) производные.

§ 4. Обратная функция. Параметрическое задание функции

Задание y как функции x означает, что каждому x соответствует определенное значение y. Часто при этом и обратно каждому значению y соответствует свое x: так, если $y=8x^3-9$, то $x^3=\frac{y+9}{8}$ и $x=\sqrt[3]{\frac{y+9}{8}}=\frac{\sqrt[3]{y+9}}{2}$. В этих случаях задание функции $y=y(x)$ дает также функциональную зависимость $x(y)$. Эту функциональную зависимость[1], как уже говорилось выше (см. § 1.6), называют обратной функцией для функции $y(x)$; в свою очередь зависимость $y(x)$ будет обратной для функции $x(y)$.

Во многих случаях обратная функция имеет более простой вид, чем исходная: так, функция $y=\sqrt[3]{x-1}$ содержит кубический корень, а обратная ей функция $x=y^3+1$ является многочленом и проще функции $y(x)$. В подобной ситуации легче найти производную обратной функции $\frac{dx}{dy}$, чем производную $\frac{dy}{dx}$ исходной, «прямой», функции. Нельзя ли в этом случае производную прямой функции как-то выразить через производную обратной функции?

Для функции $y(x)$ имеем

$$dy=\frac{dy}{dx}\,dx=y'(x)\,dx. \qquad (4.4.1)$$

[1] В общем случае зависимость $x=x(y)$, обратная данной функции $y=y(x)$, будет *многозначной* (см. § 1.6, ср. § 11).

Соотношение (1) позволяет найти и *производную $x'(y)$ обратной функции $x(y)$*:

$$x'(y)=\frac{dx}{dy}=\frac{1}{y'(x)}. \qquad (4.4.2)$$

(Конечно, (2) снова демонстрирует, что с дифференциалами можно оперировать как с числами: $\frac{dx}{dy}=1\Big/\frac{dy}{dx}$.) В правой части (2) стоит выражение, записанное в виде функции от x. Но если известна зависимость $x=x(y)$, то это выражение можно представить и как функцию от y, а именно как $\frac{1}{y'(x(y))}$.

Поясним сказанное примерами. Первый из известных нам примеров производных (линейная функция) слишком прост.

Начнем со второго примера:

$$y=x^2, \quad \frac{dy}{dx}=y'(x)=2x,$$

$$\frac{dx}{dy}=\frac{1}{y'(x)}=\frac{1}{2x}. \qquad (4.4.3)$$

Подставляем в выражение (3) обратную функцию $x=\sqrt{y}$; получим

$$\frac{dx}{dy}=\frac{d\sqrt{y}}{dy}=\frac{1}{2x}=\frac{1}{2\sqrt{y}}. \qquad (4.4.4)$$

Раньше, в гл. 2, этот результат (производная функции $y=\sqrt{x}$) был получен более сложным путем (см. упр. 2.4.2).

Вот еще один пример:

$$y=x^3+1, \quad \frac{dy}{dx}=y'(x)=3x^2,$$

$$\frac{dx}{dy}=\frac{1}{y'(x)}=\frac{1}{3x^2},$$

$$\frac{dx}{dy}=\frac{d\left(\sqrt[3]{y-1}\right)}{dy}=\frac{1}{3x^2}=$$

$$=\frac{1}{3\sqrt[3]{(y-1)^2}}=\frac{1}{3}(y-1)^{-2/3}$$

(ср. § 5).

Параметрическое задание функции (см. § 1.8) можно рассматривать как частный случай сложной функции. В самом деле, если дано

$$x=f(t), \quad y=g(t), \qquad (4.4.5)$$

то первое из равенств (5) можно рассматривать как уравнение, решив ко-

торое относительно неизвестного t, мы найдем зависимость $t=t(x)$; подставляя это выражение $t(x)$ во второе уравнение, получим $y=g(t)=g(t(x))$. Следовательно,

$$\frac{dy}{dx}=\frac{dy}{dt}\frac{dt}{dx}.$$

Но для применения этой формулы необязательно выражать t как функцию x (если бы мы это сделали, то избавились бы от параметра, что не всегда возможно). Достаточно знать функцию $x=f(t)$. Эта функция является обратной по отношению к функции $t(x)$.

Значит,

$$\frac{dt}{dx}=\frac{1}{\frac{dx}{dt}}.$$

Таким образом,

$$\frac{dy}{dx}=\frac{\frac{dy}{dt}}{\frac{dx}{dt}}. \qquad (4.4.6)$$

Эта формула дает еще один пример того, что с дифференциалами можно обращаться как с обычными алгебраическими величинами: величина dt в правой части равенства (6) сокращается.

Приведем пример. Пусть

$$x=t^2-t, \quad y=t^2+t. \qquad (4.4.7)$$

Тогда

$$\frac{dx}{dt}=2t-1, \quad \frac{dy}{dt}=2t+1,$$

$$\frac{dy}{dx}=\frac{2t+1}{2t-1}.$$

При построении графика параметрически заданной функции $x=x(t)$, $y=y(t)$ мы составляем таблицу, где указываем отвечающие выбранным значениям t величины x и y, задающие точку $M=M(x, y)$ графика. Но удобно попутно вычислять при выбранном t также и производные $\frac{dx}{dt}$ и $\frac{dy}{dt}$; тогда их отношение $\frac{dy}{dt}\Big/\frac{dx}{dt}$ будет указывать наклон касательной к графику в той же точке $M(x, y)$.

Из (2) и (3.5) также следует формула для второй производной функции $x=x(y)$, обратной известной функции $y=y(x)$:

$$\frac{d^2x}{dy^2}=\frac{d}{dy}\left(\frac{dx}{dy}\right)=\frac{d}{dy}\left(\frac{1}{y'(x)}\right)=$$

$$=-\frac{\frac{d}{dy}(y'(x))}{(y'(x))^2}=-\frac{\frac{d}{dx}(y'(x))\frac{dx}{dy}}{(y'(x))^2}=$$

$$=-\frac{y''(x)\left[1\Big/\left(\frac{dy}{dx}\right)\right]}{(y'(x))^2}=-\frac{y''(x)(1/y'(x))}{(y'(x))^2}.$$

Таким образом, окончательно

$$x''(y)=-\frac{y''(x)}{(y'(x))^3}. \qquad (4.4.8)$$

Так, например если $y=x^2$, $x=\sqrt{y}$, то $y'=2x$, $y''=2$, и, следовательно,

$$x''(y)=\frac{d^2(\sqrt{y})}{dy^2}=-\frac{2}{(2x)^3}=-\frac{1}{4x^3}=-\frac{1}{4y^{3/2}}.$$

Аналогично из (6) нетрудно вывести формулу для второй производной функции $y=y(x)$, заданной параметрическими уравнениями $x=x(t)$, $y=y(t)$:

$$\frac{d^2y}{dx^2}=\frac{d}{dx}\left(\frac{dy}{dx}\right), \quad \text{где } x=x(t),$$

$$\frac{dy}{dx}=\frac{y'(t)}{x'(t)};$$

следовательно,

$$\frac{d^2y}{dx^2}=\frac{\frac{d}{dt}\left(\frac{y'(t)}{x'(t)}\right)}{x'(t)}=\frac{\frac{y''(t)x'(t)-y'(t)x''(t)}{(x'(t))^2}}{x'(t)},$$

т. е.

$$\frac{d^2y}{dx^2}=\frac{y''(t)x'(t)-y'(t)x''(t)}{(x'(t))^3}. \qquad (4.4.9)$$

Так, в случае той же функции (7) имеем $\frac{d^2x}{dt^2}=2$, $\frac{d^2y}{dt^2}=2$ и

$$\frac{d^2y}{dx^2}=\frac{2(2t-1)-(2t+1)\cdot 2}{(2t-1)^3}=-\frac{4}{(2t-1)^3}.$$

Упражнения

4.4.1. Найдите первую производную функций: а) $y=\sqrt[3]{2x+1}$; б) $y=\sqrt{\frac{x-1}{x+1}}$; в) $y=\frac{1}{\sqrt[4]{x}}$.

4.4.2. $x=a\cos t$, $y=b\sin t$; чему равна производная $\frac{dy}{dx}$?

4.4.3. Найдите вторую производную $\frac{d^2y}{dx^2}$ функций: а) $y=\sqrt{\frac{x-1}{x+1}}$; б) $y=\frac{1}{\sqrt[4]{x}}$; в) $x=a\cos t$, $y=b\sin t$.

§ 5. Степенная функция

Рассмотрим степенную функцию

$$y = x^n, \qquad (4.5.1)$$

где n — некоторое число. При n целом положительном x^n есть произведение n одинаковых множителей:

$$y = \underbrace{x \cdot x \cdot \ \ldots \ \cdot x}_{n \text{ раз}},$$

и, значит (по формуле типа (2.7)),

$$\frac{dy}{dx} = \underbrace{1 \cdot x^{n-1} + 1 \cdot x^{n-1} + \ldots + 1 \cdot x^{n-1}}_{n \text{ слагаемых}},$$

откуда

$$\frac{dy}{dx} = nx^{n-1}. \qquad (4.5.2)$$

Покажем, что эта формула справедлива и при любом n (дробном, отрицательном)[1].

При дробном n положим $n = \frac{m}{p}$, где числа m и p — целые:

$$y = x^{m/p}, \qquad (4.5.3а)$$

или

$$y^p = x^m. \qquad (4.5.3б)$$

Выражение y^p в левой части равенства (3б) является сложной функцией от x, так как y зависит от x. Поэтому, вычисляя производную от обеих частей равенства (3б) и учитывая (3.3а), получим

$$\frac{d}{dx}(y^p) = \frac{d}{dx}(x^m), \text{ или}$$

$$py^{p-1}\frac{dy}{dx} = mx^{m-1}.$$

Отсюда

$$\frac{dy}{dx} = \frac{m}{p}\frac{x^{m-1}}{y^{p-1}} = \frac{m}{p}\frac{x^{m-1}}{(x^{m/p})^{p-1}} = \frac{m}{p}x^{\frac{m}{p}-1}.$$

Учитывая, что $\frac{m}{p} = n$, имеем окончательно

$$\frac{dy}{dx} = nx^{n-1}.$$

Если показатель степени отрицателен, т. е. $n = -k$, где k — положительное число, то

$$y = x^n = x^{-k} = \frac{1}{x^k}.$$

Используем снова (3.3а) $\left(\text{или, точнее, (3.5); здесь } y = \frac{1}{f},\ f = x^k\right)$, найдем

$$\frac{dy}{dx} = -\frac{1}{f^2}\frac{df}{dx} = -\frac{1}{x^{2k}}kx^{k-1} = -kx^{-k-1}.$$

Вспоминая, что $k = -n$, получим и для отрицательного n

$$\frac{dy}{dx} = \frac{dx^n}{dx} = nx^{n-1}.$$

Таким образом, формула производной от степени применима при любом рациональном показателе n. Она распространяется и на случай иррационального показателя степени[2].

Формула (2) имеет важнейшее значение. Из нее, в частности, следует, что при малых по абсолютной величине r

$$(1 + r)^n \approx 1 + nr. \qquad (4.5.4)$$

В самом деле, если $x = 1$, $\Delta x = r$, $y = x^n$, то $\Delta y = (x + \Delta x)^n - x^n = (1 + r)^n - 1$. С другой стороны, $dy = y'(x)\Delta x = n \cdot 1^{n-1} \cdot r = nr$. Теперь (4) вытекает из того, что (при малых $\Delta x = r$) приращение Δy близко к дифференциалу dy.

Формулу (2) полезно записать еще и в другом виде:

$$\text{если } y = cx^n, \text{ то } \frac{dy}{dx} = n\frac{y}{x}. \qquad (4.5.5)$$

Этот результат надо глубоко прочувствовать. При положительных n степенная функция обладает тем очевидным свойством, что при $x = 0$ также и $y = 0$. Кривую $y = cx^n$ при данном $n > 0$ можно провести через любую точку (x_0, y_0) — достаточно выбрать $c = \frac{y_0}{x_0^n}$. Пусть кривая проходит через начало координат и через точку (x_0, y_0). Найдем среднее значение производной на участке кривой от начала координат до точки (x_0, y_0).

Согласно определению среднего (см. § 7.7)

$$\overline{y'} = \frac{\int_0^{x_0} y'(x)\,dx}{x_0 - 0} = \frac{1}{x_0}\int_0^{x_0} y'(x)\,dx,$$

[1] Легче всего видеть, что она справедлива при $n = 0$, когда $y = \text{const}$ и $y' = 0$.

[2] Функция $y = x^n$, где n — иррационально, определяется условием $x^n = \lim x^r$, где r — рационально и $\lim r = n$; поэтому если формула (2) справедлива для всех рациональных показателей степени n, то она неизбежно будет выполняться и для иррациональных показателей степени.

откуда в силу (3.4.9)

$$\overline{y'} = \frac{y(x_0) - y(0)}{x_0} = \frac{y(x_0)}{x_0} = \frac{y_0}{x_0}.$$

В самом деле, при изменении x от 0 до x_0 величина y растет от 0 до y_0. Значит, средняя скорость роста y (т. е. среднее значение производной) равна $\frac{y_0}{x_0}$, что очевидно и без интегралов!

Как видно из (5), значение производной в точке (x_0, y_0) равно n-кратному среднему значению производной (где n — показатель степени). На рис. 1 показано несколько кривых,

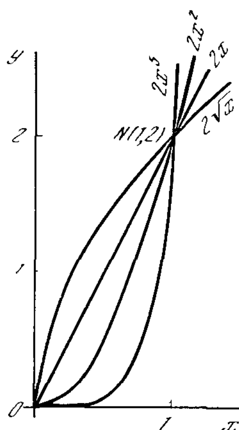


Рис. 4.5.1

отвечающих различным значениям n ($n = 1/2, 1, 2, 5$) и проходящих через одну и ту же точку N (x_0, y_0), а значит, имеющих одинаковую среднюю производную на отрезке $[0, x_0]$. Наглядно видно, что чем больше n, тем больше производная в точке N (тем круче растет кривая).

Вернемся еще раз к формуле (5). Из нее вытекает, что $dy = n\frac{y}{x}dx$, и поэтому для малых приращений Δx

$$\Delta y \approx n\frac{y}{x}\Delta x. \tag{4.5.6}$$

Для $\Delta x = 0{,}01x$, т. е. для изменения аргумента на 1%, соотношение (6) можно считать достаточно точным. Из (6) следует

$$\Delta y \approx n\frac{y}{x}\cdot 0{,}01x,$$

или

$$\Delta y \approx n \cdot 0{,}01y.$$

Таким образом, при изменении аргумента на 1% значение степенной функции (1) меняется на n%.

Упражнения

4.5.1. Найдите производные функций: а) $y = x^5 - 3x^4 + x^3 + 7x^2 - 2x + 5$; б) $y = (x^3 + x + 1)^2$; в) $y = (x^2 - x + 1)^4$; г) $y = (3x^2 - 1)^{10}$; д) $y = \sqrt{x^2 - 1}$; е) $y = \sqrt[5]{x^2}$.

4.5.2. Получите результат (2) (для целого положительного n) с помощью формулы бинома Ньютона. [Указание. Используйте биномиальную формулу (см. § 6.4) для вычисления $\Delta x = (x + \Delta x)^n - x^n$.]

4.5.3. Найдите значения $y(9)$ и $y(11)$, если $y(10) = 5$ и: а) $y \sim \sqrt{x}$, б) $y \sim \frac{1}{x}$, в) $y \sim x^2$, где знак $\sim$ означает *пропорционально*. Задачу решите в уме, без выкладок. Сравните ответ с точным.

§ 6. Производные алгебраических функций

Совокупность правил § 1—5 позволяет найти производную любой алгебраической функции, образованной из независимой переменной x с помощью операций сложения и вычитания, умножения и деления, а также операции возведения в (произвольную) степень, в том числе и операции возведения в дробную степень, т. е. извлечения корня.

Покажем на примере, как это удобнее всего делать практически. Найдем производную функции

$$f(x) = x\sqrt[3]{x^2 - 1}.$$

Ответ следует писать сразу, т. е. не вводя каких-либо новых промежуточных обозначений (вроде $\sqrt[3]{x^2 - 1} = u$). Производную берут как бы отдельно «по каждому месту», где стоит x, приговаривая для памяти примерно следующее (буквы а, б, в. . . показывают, к каким частям написанного ниже выражения производной относятся слова): производная (а) по x, стоящему перед корнем, плюс (б) производная по $\sqrt[3]{x^2 - 1}$, умноженная на (в) производную от $\sqrt[3]{x^2 - 1}$ по $x^2 - 1$, умноженная на (г) производную от $x^2 - 1$ по x, т. е.

$$\frac{df}{dx} = \underbrace{1\cdot\sqrt[3]{x^2 - 1}}_{(а)} + \underbrace{x}_{(б)}\cdot\underbrace{\frac{1}{3}\frac{\sqrt[3]{x^2 - 1}}{x^2 - 1}}_{(в)}\underbrace{2x}_{(г)}. \tag{4.6.1}$$

Имеет смысл сразу приучаться к такому деловому способу, без лишнего чистописания, пользуясь следующими принципами:

а) правило дифференцирования сложной функции (§ 3, формулы (3.3а), (3.6));

б) если выражение составлено из нескольких функций, то его производная равна сумме производных, вычисленных в предположении, что каждый раз лишь одна из функций предполагается переменной, а остальные постоянны (§ 2, формулы (2.4), (2.7); см. также (2.1), (2.2) и (3.8)).

Формулу для производной степени удобно применять в виде

$$y = cx^n, \quad \frac{dy}{dx} = n\frac{y}{x},$$

как это сделано выше (см. выражение (в)).

Для того чтобы приобрести навык, нужно проделать 10—20 упражнений только на технику дифференцирования, безотносительно к тем физическим, геометрическим и другим задачам, в которых возникает потребность в нахождении производных.

Упражнения

Найдите производные функций:

4.6.1. $y = x^3(x^2-1)^2$. **4.6.2.** $y = x^3\sqrt{x^2+x}$. **4.6.3.** $y = x^5\sqrt[3]{x^2-1}\,(x^3-2x)^{1/5}$. **4.6.4.** $y = \left(x + \frac{1}{\sqrt{x}}\right)\sqrt{x^3-2}$. **4.6.5.** $y = x^2\sqrt{\sqrt[3]{x}+x}$. **4.6.6.** $y = \left(\sqrt[3]{x} + \frac{1}{\sqrt[3]{x}}\right)^5 x$. **4.6.7.** $y = \frac{x}{1-x^2}$. **4.6.8.** $y = \frac{x^2+x+1}{x^2-x+1}$. **4.6.9.** $y = \frac{(x-1)(x+3)}{x+1}$. **4.6.10.** $y = \frac{3x-1}{x^5}\sqrt{x^3+2}$. **4.6.11.** $y = \frac{x}{\sqrt{x^2-1}}$. **4.6.12.** $y = \frac{x}{\sqrt[3]{x+1}}$. **4.6.13.** $y = \sqrt{x^2+x\sqrt{x}}$. **4.6.14.** $y = \sqrt{x^2+\sqrt[3]{x}}$. **4.6.15.** $y = \frac{x^2-2}{(1+x^2)\sqrt{1+x^2}}$. **4.6.16.** $y = x\sqrt[3]{(2x+3)^2}$. **4.6.17.** $y = (x^3-1)\sqrt{x-1} + x\sqrt[3]{x^2-1}$. **4.6.18.** $y = \frac{x\sqrt[3]{(2x-3)^2}}{(x-1)^2}$. **4.6.19.** $y = \sqrt{\frac{x-1}{x+1}}$. **4.6.20.** $y = \frac{x^2+x+1}{x-2}\sqrt[3]{x+1}$. **4.6.21.** $y = \frac{x\sqrt{x^2-1}}{x^3+1}$. **4.6.22.** $y = \sqrt[3]{\frac{x^2+x+1}{x+1}}$. **4.6.23.** $y = x\sqrt{x^2-1}\sqrt[3]{x+\sqrt{x}}$. **4.6.24.** $y = \left(x + \frac{1}{\sqrt{x}}\right)^{1/} x^2$. **4.6.25.** $y = \frac{\sqrt[3]{x}-2x}{(1+x)^{1/3}}$.

§ 7. Показательная функция

Рассмотрим функцию

$$y = a^x,$$

где $a > 1$. Графики функции y при $a=2$ и $a=3$ изображены на рис. 1, *а*. При $x=0$ независимо от значения a имеем $y=1$.

Функция y при всех x положительна и растет с увеличением x, так что

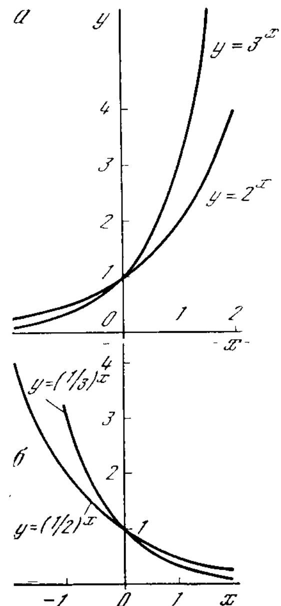


Рис. 4.7.1

и производная ее также везде положительна. При увеличении x на постоянную величину c получим

$$y(x+c) = a^{x+c} = a^c a^x = ba^x = by(x),$$

где $b = a^c$;

таким образом, y при этом умножается на постоянную величину. Поэтому если x менять последовательно, одинаковыми шагами (*в арифметической прогрессии*):

$$x = x_0,\ x_0+c,\ x_0+2c,\ \ldots,\ x_0+nc,\ \ldots,$$

то y будет принимать значения

$$y_0,\ by_0,\ b^2y_0,\ \ldots,\ b^ny_0,\ \ldots$$

Такой закон нарастания, как известно, называется *геометрической прогрессией*.

Найдем производную показательной функции[1] для $a=10$. Мы можем написать

$$\frac{d(10^x)}{dx}=\frac{10^{x+dx}-10^x}{dx}=10^x\frac{10^{dx}-1}{dx},$$

где запись dx (вместо Δx) символизирует стремление приращения к нулю.[2]

Что представляет собой величина $\frac{10^{dx}-1}{dx}$? Это есть предел отношения $\frac{10^{\Delta x}-1}{\Delta x}$ при $\Delta x\to 0$. Найдем этот предел численно, «арифметически». Пользуясь четырехзначной таблицей логарифмов (или, что еще лучше, микрокомпьютером), получаем

$$10^{0,1}\approx 1,2589,\quad \frac{10^{0,1}-1}{0,1}\approx 2,589;$$

$$10^{0,01}\approx 1,0233,\quad \frac{10^{0,01}-1}{0,01}\approx 2,33;$$

$$10^{0,001}\approx 1,0023,\quad \frac{10^{0,001}-1}{0,001}\approx 2,3.$$

Таким образом, можно заключить, что

$$\lim_{\Delta x\to 0}\frac{10^{\Delta x}-1}{\Delta x}\approx 2,3$$

(где, разумеется, стоящая справа постоянная определена нами лишь приближенно, хоть и достаточно надежно). Следовательно,

$$\frac{d}{dx}(10^x)\approx 10^x\cdot 2,3. \qquad (4.7.1)$$

Производную 10^x мы нашли, так сказать, «опытным путем», при помощи «арифметического эксперимента», приближенных вычислений. Для любой другой показательной функции задачу теперь легко свести к предыдущей. Пользуясь понятием логарифма, запишем:

$$a=10^{\lg a},\quad a^x=10^{x\lg a}. \qquad (4.7.2)$$

По правилу нахождения производной сложной функции получим

$$\frac{da^x}{dx}\approx 10^{x\lg a}\cdot 2,3\lg a=a^x\cdot 2,3\lg a. \qquad (4.7.3)$$

Замечательная особенность показательной функции заключается в том, что ее *производная прямо пропорциональна самой функции*:

$$\frac{da^x}{dx}=ka^x, \qquad (4.7.4)$$

где коэффициент пропорциональности k зависит от основания степени a. В этом главное свойство геометрической прогрессии: чем больше достигнутое значение величины, тем быстрее ее рост[3].

Если для показательной функции $y=a^x$ имеем $0<a<1$, то график функции имеет вид, изображенный на рис. 1, *б*. Здесь при увеличении x в арифметической прогрессии y у м е н ь ш а е т с я в геометрической прогрессии. Формула (3) в этом случае попрежнему применима; однако теперь величина $\lg a$ отрицательна (ибо $a<1$) и, следовательно, производная, будучи пропорциональна функции, имеет противоположный ей знак. Во второй части книги мы рассмотрим ряд задач, в которых та или иная величина уменьшается с течением времени, притом так, что скорость ее убывания пропорциональна самой оставшейся на данный момент величине:

$$\frac{dy}{dx}=-cy,\quad \text{где } c>0. \qquad (4.7.5)$$

Как видно из предыдущего, в этом случае решением задачи явится показательная функция

$$y=y_0a^x\quad (\text{где } a<1). \qquad (4.7.6)$$

Упражнения

Найдите производные функций:

4.7.1. $y=2^x$. **4.7.2.** $y=5^{x+1}$. **4.7.3.** $y=\left(\frac{1}{2}\right)^x$. **4.7.4.** $y=10^{\sqrt{x}}$. **4.7.5.** $y=2^{x^2}$. **4.7.6.** $y=2^{x+\frac{1}{x}}$.

---

[1] $a=10$ взято для облегчения вычислений с помощью таблицы десятичных логарифмов. Читатель, пользующийся компьютером с кнопкой y^x, может вместо этого выписать, например, $2^{0,1}$, $2^{0,01}$, . . . или $3^{0,1}$, $3^{0,01}$, . . . и соответствующие этим числам отношения $(2^{0,1}-1)/0,1$ и т. д.

[2] Именно это обстоятельство позволяет нам смело писать $\frac{d(10^x)}{dx}=\frac{10^{x+dx}-10^x}{dx}$ (т о ч н о е равенство!) вместо $\frac{d(10^x)}{dx}=\frac{10^{x+\Delta x}-10^x}{\Delta x}$ (п р и б л и ж е н н о е равенство).

[3] Свойство геометрической прогрессии — необычайно быстрое ее возрастание — любимая тема многих популярных книг по математике, например «Занимательной алгебры» и «Живой математики» Я. И. Перельмана.

§ 8. Число e

Найдем такое значение основания a, для которого формула производной показательной функции $y=a^x$ имела бы наиболее простой вид. А именно потребуем, чтобы в выражении (7.4) коэффициент пропорциональности k равнялся единице (в этом случае его можно вовсе не писать). Обозначим искомое основание степени буквой e. Таким образом, по определению числа e, имеем

$$\frac{de^x}{dx}=e^x. \qquad (4.8.1)$$

Число e легко найти с помощью формулы (7.3):

$$2{,}3\lg e\approx 1, \quad \lg e\approx\frac{1}{2{,}3}\approx 0{,}43,$$

откуда по таблице логарифмов

$$e\approx 2{,}7.$$

Однако такой практический подход не соответствует историческому ходу развития науки и принципиально не удовлетворителен. Мы пользовались числами, заимствованными из таблицы логарифмов, не задумываясь над тем, как эти числа были получены [1].

Найдем число e, основываясь только на формуле (1). В силу общих свойств показательной функции $e^0=1$. Рассмотрим функцию $y=e^x$. Мы имеем $y(0)=1$ и согласно (1) $y'(0)=1$.

Возьмем малое $\Delta x=r$ и подсчитаем приращение функции $y=e^x$ при переходе от $x=0$ к $x=r$. Как мы знаем, при этом можно писать $\Delta y\approx y'\,\Delta x$. Поэтому

$$\Delta y\approx 1\cdot\Delta x=r, \quad y(x)=y(0)+\Delta y,$$

откуда

$$e^r\approx 1+r. \qquad (4.8.2)$$

Это равенство надо понимать в том же смысле, в каком понималось выше равенство

$$y(x+\Delta x)\approx y(x)+y'(x)\Delta x,$$

т. е. в равенстве (2) опущены члены второго и высших порядков, т. е. члены, пропорциональные r^2, r^3 и т. д., и сохранен лишь член первого порядка малости, пропорциональный r. При этом важно, что коэффициент при r точно равен единице — это следует из свойства числа e, записанного формулой (1) и являющегося определением e.

Представим себе число r в виде дроби с большим знаменателем: $r=\frac{1}{n}$; если[2] $r\ll 1$, то $n\gg 1$. Тогда из (2) следует

$$e^{1/n}\approx 1+\frac{1}{n}, \quad \text{откуда } e\approx\left(1+\frac{1}{n}\right)^n.$$

Так как формула (2), пренебрегающая членами с r^2 и др., тем точнее, чем меньше r, то последнее выражение тем точнее, чем больше n. Таким образом, мы получаем второе строгое определение числа e:

$$e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \qquad (4.8.3)$$

(читается: e есть предел выражения $\left(1+\frac{1}{n}\right)^n$ при n, стремящемся к бесконечности[3]).

Аналогично при фиксированном k и достаточно малом r (таком, что $kr\ll 1$) имеем

$$e^{kr}\approx 1+kr,$$

---

[1] Точность с которой можно таким образом определить число e, также весьма невелика: вряд ли, определяя производную 10^x при помощи четырехзначной таблицы логарифмов, Вы сможете верно найти сотые доли в десятичном представлении числа e. В действительности $\frac{10^{dx}-1}{dx}\approx 2{,}302585$, но, конечно, это значение с помощью таблицы логарифмов получить не удается. Еще менее соответствует возможностям школьных таблиц логарифмов легко запоминаемое, но практически достаточно бесцельное приближенное значение числа e с 15 десятичными знаками; $e\approx$ $\approx 2{,}718281828459045$. (Заметим, что число e не выражается *периодической* десятичной дробью.)

По поводу формул (и методов), позволяющих вычислить e с очень большой степенью точности, см. § 6.1, 6.2.

[2] Запись $r\ll 1$ означает, что число r *значительно меньше* 1; аналогично запись $n\gg 1$ означает, что n — большое число (n *значительно больше* 1).

[3] Разумеется, с точки зрения ригориста-математика, уже переход «из $e^{1/n}\approx 1+\frac{1}{n}$ следует $e\approx\left(1+\frac{1}{n}\right)^n$» не строг, так как обе части приближенного равенства (верного лишь при $n\gg 1$) не всегда можно возводить в большую степень n. Таким образом, наше рассуждение не доказывает определение (3), а всего лишь поясняет его. Тем не менее мы надеемся, что читателю это рассуждение покажется убедительным.

откуда, положив $r=\frac{1}{n}$, где n очень велико, получаем

$$e^k \approx \left(1+\frac{k}{n}\right)^n,$$

или, точнее,

$$e^k = \lim_{n\to\infty}\left(1+\frac{k}{n}\right)^n. \qquad (4.8.3а)$$

Не надо бояться слов «предел», «бесконечность». Практически например, $\left(1+\frac{1}{100}\right)^{100} \approx 2{,}705$ уже довольно мало отличается от точного значения e. Советуем читателю самому найти $\left(1+\frac{1}{8}\right)^8$.

Мы видели, что $e^r \approx 1+r$ при малом r, и это приближенное равенство[4] тем точнее, чем меньше r. Проверим сказанное на конкретных примерах.

Составим следующую таблицу[5]:

r	$1+r$	e^r	r	$1+r$	e^r
−0,5	0,5	0,6065	0,01	1,01	1,0101
−0,4	0,6	0,6703	0,1	1,1	1,1052
−0,3	0,7	0,7408	0,2	1,2	1,2214
−0,2	0,8	0,8187	0,3	1,3	1,3499
−0,1	0,9	0,9048	0,4	1,4	1,4918
−0,01	0,99	0,9900	0,5	1,5	1,6487
0	1	1			

Видно, что даже при $r=\pm 0{,}3$ ошибка в (2) не превышает 6%. Физику (или инженеру) полезно помнить не только то, что $e\approx 2{,}72$, но и приближенные значения: $e^2\approx 7{,}4$, $e^3\approx 20$, $e^4\approx 55$, $e^5\approx 150$. Краткие таблицы значений функций e^x и e^{-x} даны в Приложении IV к книге.

При помощи числа e упрощается решение задач на геометрические прогрессии и сложные проценты. Пусть, скажем, требуется указать, во сколько раз вырастет производство за 50 лет при ежегодном росте на 2%. Нам нужно вычислить $1{,}02^{50}$. Применение числа e заключается в том, что, пользуясь соотношением (2), мы приближенно заменяем 1,02 на $e^{0{,}02}$, в силу чего $1{,}02^{50} \approx e^{0{,}02\cdot 50} = e \approx 2{,}72$. Общая формула имеет вид

$$(1+r)^m \approx e^{mr}, \text{ где } r \ll 1. \qquad (4.8.4)$$

[4] Для функции $y=e^x$ составлены подробные таблицы.

[5] Значения e^x взяты из четырехзначных таблиц Брадиса.

Для применимости этой формулы достаточно, чтобы было мало́ r; m и mr могут и не быть малыми. Если mr тоже мало́, то $e^{mr}\approx 1+mr$, так что мы получаем известную ранее формулу

$$(1+r)^m \approx 1+mr \qquad (4.8.5)$$

(см. формулу (5.4)). Однако при больших mr формулой (5) пользоваться уже нельзя, тогда как выражение (4) остается справедливым и в этом случае (если только $r \ll 1$). Так, для примера, приведенного выше, точное значение числа $1{,}02^{50}$ (с тремя знаками после запятой) равно 2,693, а формула (4) дает $1{,}02^{50}\approx e^1 \approx 2{,}72$; по формуле же (5) мы получим $(1+0{,}2)^{50}\approx 1+50\cdot 0{,}2=2$. Расчет с помощью числа e (формула (4)) дал ошибку около 1%, тогда как расчет по формуле (5) здесь дал ошибку порядка 25%.

В общем виде оценка точности формулы (4) дана в § 6.1 (см. упр. 6.1.4). Оказывается, что относительная погрешность этой формулы имеет величину порядка mr^2, так что формула (4) применима при $mr^2 \ll 1$, а формула (5), как мы знаем, — лишь при $mr \ll 1$. Но так как (увы!) условие $mr \ll 1$ при малых r является гораздо более сильным, чем условие $mr^2 \ll 1$, то существуют значения m и r, при которых можно пользоваться приближенным равенством (4), но не равенством (5). Например, в рассмотренном нами случае $r=0{,}02$, $m=50$ имеем: $mr=1$, а $mr^2=0{,}02$, т. е. mr^2 мало́, а величину mr малой считать нельзя. Так же, скажем, при $r=0{,}001$, $m=10\,000$ имеем $mr=10$ и условие $mr \ll 1$ никак не выполняется; однако $mr^2=0{,}01$, так что здесь вполне можно считать выполненным условие $mr^2 \ll 1$ (ср. упр. 2).

В соответствии с первоначальным определением числа e (формула (1)) производные от показательных функций имеют особенно простой вид, когда основанием степени является число e. Эти производные удобно выражаются через саму функцию. Приведем ряд формул:

$$y=e^x, \quad \frac{dy}{dx}=e^x=y;$$

$$y=Ce^x, \quad \frac{dy}{dx}=C\frac{de^x}{dx}=Ce^x=y;$$

$$y=Ce^{kx},$$

$$\frac{dy}{dx}=C\frac{de^{kx}}{dx}=Ce^{kx}\frac{d(kx)}{dx}=ky;$$

$$y = e^{m(x)},$$

$$\frac{dy}{dx} = \frac{de^{m(x)}}{dx} = e^{m(x)} \frac{dm(x)}{dx} = y \frac{dm(x)}{dx};$$

$$y = f(x) e^{m(x)},$$

$$\frac{dy}{dx} = f'(x) e^{m(x)} + f(x) e^{m(x)} m'(x) = y \left(\frac{f'(x)}{f(x)} + m'(x) \right).$$

Для показательной функции $y=e^x$ (с основанием степени e) часто используют запись $y=\exp(x)$ (от латинского

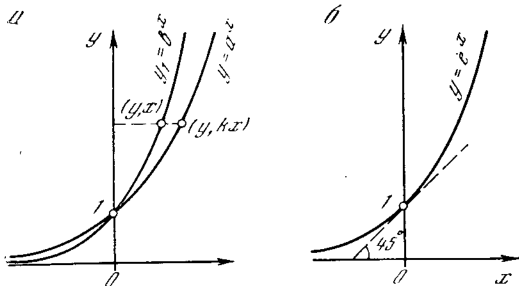


Рис. 4.8.1

слова exponent — показатель; сама функция e^x называется *экспоненциальной* функцией, или зависимостью, а запись $\exp x$ читается как «экспонента икс»). Запись с помощью символа exp особенно удобна в тех случаях, когда аргументом функции служит сложное выражение: ясно, что удобнее написать

$$\exp\left(\frac{7t^2+24t}{t^3+5}\right)^3, \text{ чем } e^{\left(\frac{7t^2+24t}{t^3+5}\right)^3}.$$

Число e можно также определить геометрически. Показательные функции $y=a^x$ и $y_1=b^x$ связаны простой зависимостью

$$b^x = a^{kx}, \quad \text{где } k = \log_a b \tag{4.8.6}$$

(этот факт уже использовался в формуле (7.2)). Из (6) следует, что график функции y_1 получается из графика функции y сжатием в k раз в направлении оси x (рис. 1, *а*, ср. § 1.7).

Ясно, что все графики функций $y=a^x$, отвечающие различным a, проходят через одну и ту же точку (0,1) (ибо при любом a имеем $a^0=1$); в этой точке они пересекают ось y под разными углами. Самой простой можно счесть ту из показательных функций, график которой (точнее, касательная к графику которой) в точке (0,1) образует с осью y угол 45° (рис. 1, *б*). Из формулы (1) вытекает, что этому условию удовлетворяет как раз функция e^x (вспомните геометрический смысл производной!).

В ы в о д. Итак, можно дать четыре различных определения числа e:

1) условием $(e^x)'=e^x$;

2) условием $e^r \approx 1+r$, выполняющимся при $r \ll 1$;

3) как предел выражения $\left(1+\frac{1}{n}\right)^n$ при $n \to \infty$;

4) как основание степенной функции $y=e^x$, график которой пересекает ось y под углом 45°.

У числа e есть и другие замечательные определения и свойства; о некоторых из них будет рассказано в гл. 6.

Для закрепления этого очень важного раздела читатель должен, отложив книгу, сам показать, как из каждого определения следуют три другие, рассматриваемые как свойства числа e.

Заметим, наконец, что в равенстве $y=e^x$ величины x и y — это, безусловно, (безразмерные) числа, а не именованные величины: ведь запись e^x, где x — это 100 м или 5 ч, не имеет никакого смысла. Соответственно этому в естественнонаучных законах экспоненциальная зависимость между размерными величинами y и x обязательно имеет вид $y=l_2 e^{x/l_1}$, где l_1 и l_2 — единицы измерения x и y. При этом переход к новой единице $l_1'=cl_1$ измерения величин x приводит к замене исходной зависимости $y=e^x$ (по существу, равносильной ей) зависимостью $y=e^{cx'}$ $\left(\text{где } x'=\frac{x}{c}\text{ — та же величина } x, \text{ но измеренная в новых единицах } l_1'\right)$. Аналогично этому замена $l_2'=\frac{1}{d}l_2$ единиц измерения величин y приводит к замене формулы $y=e^x$ на $y'=de^x$ (где y' уже измеряется в единицах l_2'). Соответственно этому экспоненциальными обычно называют все зависимости типа $y=ae^{kx}$, где a и k могут быть какими угодно: ведь в большинстве случаев переход к другим a и k означает лишь изменение системы единиц, так что физический смысл имеет лишь сам факт степенной зависимости и знаки чисел a и k (в частности, условию $k>0$ отвечает растущая функция $y=e^{kx}$, а условию $k<0$ — убывающая).

Упражнения

4.8.1. Найдите производные следующих функций:

а) $y=e^{-x}$; б) $y=5e^{x}-e^{3x}$; в) $y=e^{x^2}$; г) $y=e^{\sqrt{x}}$; д) $y=e^{x^3-3x+1}$.

4.8.2. Сумма увеличивается на 0,1% ежемесячно; во сколько раз вырастет она за 1000 лет? [Указание. Используйте формулу (4).]

§ 9. Логарифмы

Как известно, *логарифмом* числа h по основанию a называется та степень f, в которую надо возвести a, чтобы получить h:

$$f=\log_a h \text{ означает, что } h=a^f. \quad (4.9.1)$$

Таким образом, логарифмическая функция $y=\log_a x$ обратна (в смысле § 1.6) показательной функции $y=a^x$.

График логарифмической функции $y=\log_a x$ (где $a>1$) изображен на рис. 1. Отметим, что при $x=1$ (независимо от значения a) имеем $y=0$, а при $x>1$ величина $y>0$ (при $x<1$ соответственно $y<0$). Вся кривая $y=\log_a x$ расположена правее оси ординат: так как результат возведения положительного числа a в любую степень будет положительным, то отрицательные числа логарифмов не имеют [1].

Из рис. 1 видно, что производная функции $y=\log_a x$ (при $a>1$) при всех значениях x положительна, ибо функция y при всех x растет. Заметим также, что скорость роста функции $y=\log_a x$ с увеличением x уменьшается, т. е. производная функции при этом убывает (ниже мы строго докажем это).

Выведем формулу, связывающую логарифмы одного и того же числа по разным основаниям. Пусть

$$f=\log_a h, \quad a^f=h. \quad (4.9.2)$$

Прологарифмируем обе части второго из равенств (2) по основанию b:

$$f\log_b a=\log_b h, \text{ откуда } f=\frac{\log_b h}{\log_b a}.$$

Принимая во внимание (2), получаем

$$\log_a h=\frac{\log_b h}{\log_b a}. \quad (4.9.3)$$

Равенство (3) можно также переписать так:

$$\log_b x=k\log_a x, \text{ где } k=\log_b a. \quad (4.9.3а)$$

[1] Ср., впрочем, § 14.3.

Число $k=\log_b a$ называется *модулем перехода от логарифмов по основанию a к логарифмам по основанию b*. Из (3а) следует, что график функции $y=\log_b x$ получается из графика функции $y=\log_a x$ растяжением в k раз вдоль оси y.

Логарифмы по основанию e называются *натуральными* (т. е. природными, естественными) *логарифмами* и обозначаются $\ln x$. «Естественность» именно этой системы логарифмов связана с тем, что в ряде отношений они оказываются

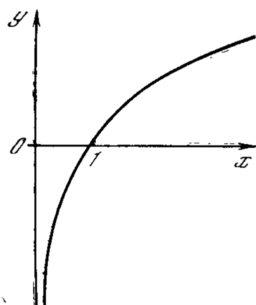


Рис. 4.9.1

простейшими: так, например, для функции $y=\ln x$ проще всего определить производную (скорость роста). Эти замечательные свойства натуральных логарифмов, с которыми мы неоднократно встретимся ниже, привели к тому, что оба создателя учения о логарифмах — шотландский дворянин Джон Непер (1550—1617) и швейцарский часовщик Йобст Бюрги (1552—1632), — почти одновременно и независимо их открывшие, рассматривали именно этот тип логарифмов (или весьма близкие к ним). *Десятичные логарифмы* впервые рассмотрел по предложению Непера его друг и почитатель лондонский профессор Генри Бригс (1561—1630) [2].

Натуральные логарифмы можно охарактеризовать тем, что график функции $y=\ln x$ пересекает ось абсцисс (в точке (1,0)) под углом 45° — ведь этот график получается из графика (обратной натуральному логарифму) функции $y=e^x$ отражением от биссектрисы координатного угла (см. § 1.6), — а график экспоненты $y=e^x$ пересекает под углом 45° ось ординат (в точке (1,0); см. рис. 8.1, *б*).

[2] В старину десятичные логарифмы зачастую называли *бригсовыми*, а натуральные — *неперовыми*.

Найдем теперь производную натурального логарифма. Рассмотрим величину $d\ln x = \ln(x+dx) - \ln x$. Из известной формулы $\ln a - \ln b = \ln\frac{a}{b}$ следует

$$d\ln x = \ln\frac{x+dx}{x} = \ln\left(1+\frac{dx}{x}\right). \qquad (4.9.4)$$

Но мы уже знаем, что при малых r

$$e^r \approx 1 + r$$

(см. формулу (8.2)). Возьмем логарифмы обеих частей последнего равенства; поскольку $\ln e^r = r$, получим очень важное соотношение

$$\ln(1+r) \approx r \qquad (4.9.5)$$

(приближенное равенство (5) имеет смысл, аналогичный (8.2); оно верно при $r \ll 1$). Из (4) и (5) следует

$$d\ln x = \ln\left(1+\frac{dx}{x}\right) = \frac{dx}{x}$$

(запись dx, а не Δx, позволяет нам считать последнее равенство точным!), поэтому

$$\frac{d\ln x}{dx} = \frac{1}{x}. \qquad (4.9.6)$$

Когда x меняется в геометрической прогрессии, $\ln x$ меняется в арифметической прогрессии:

если $x = a, ab, ab^2, ab^3, \ldots$, то $\ln x = \ln a, \ln a + c, \ln a + 2c, \ln a + 3c, \ldots$, где $c = \ln b$.

Поэтому чем больше x, тем медленнее растет $\ln x$ и тем меньше производная функции $\ln x$, что и отражает формула (6).

Производную натурального логарифма можно найти также, пользуясь тем, что логарифм и экспонента — взаимно-обратные функции. Запишем:

$$y = \ln x, \quad x = e^y, \quad x' = \frac{dx}{dy} = \frac{d(e^y)}{dy} = e^y,$$

$$\frac{dy}{dx} = \frac{1}{e^y} = \frac{1}{x}.$$

Теперь, пользуясь (3а), можно вычислить производную логарифма и по любому основанию a. Пусть $y = \log_a x$; тогда (см. (3) и (3а))

$$y = \frac{\ln x}{\ln a}, \qquad \frac{dy}{dx} = \frac{1}{\ln a}\frac{d\ln x}{dx} = \frac{1}{\ln a}\frac{1}{x}. \qquad (4.9.7)$$

Заменив в равенстве (3) a на e, а b и h на a, получим $\ln a = \frac{1}{\log_a e}$, что позволяет переписать (7) так:

$$\frac{d\log_a x}{dx} = \frac{\log_a e}{x}. \qquad (4.9.7a)$$

Из формул (6), (7) и (7а) самой простой является формула (6), верная именно для натуральных логарифмов. Для грубых расчетов в уме полезно запомнить, что $\ln 2 \approx 0{,}69$; $\ln 3 \approx 1{,}1$; $\ln 10 \approx 2{,}3 \approx \frac{1}{0{,}434}$. Краткая таблица натуральных логарифмов дана в Приложении IV к книге.

Если под знаком логарифма стоит какая-нибудь функция $f(x)$, то производную находим по правилу дифференцирования сложных функций (см. § 3):

$$\frac{d\ln f(x)}{dx} = \frac{1}{f(x)}\frac{df(x)}{dx}. \qquad (4.9.8)$$

Производную $(\ln f(x))'$ логарифма функции $f(x)$ часто называют *логарифмической производной* $f(x)$; согласно (8) она равна частному производной $f'(x)$ и самой исходной функции $f(x)$. Приведем один пример, демонстрирующий пользу понятия логарифмической производной.

Формула (8) дает возможность находить производные функций вида $f(x)^{h(x)}$, т. е. функций, содержащих переменную и в основании и в показателе степени. Пусть

$$y = f(x)^{h(x)}. \qquad (4.9.9)$$

Прологарифмируем (9) (логарифмы можно брать по любому основанию; возьмем натуральные логарифмы):

$$\ln y = h(x)\ln f(x). \qquad (4.9.10)$$

Продифференцируем теперь обе части равенства (10) (т. е. найдем производные обеих частей); при этом учтем, что $\ln y$ есть сложная функция x (так же как $\ln f(x)$):

$$\frac{1}{y}y' = h'(x)\ln f(x) + h(x)\frac{f'(x)}{f(x)},$$

откуда

$$y' = y\left[h'(x)\ln f(x) + h(x)\frac{f'(x)}{f(x)}\right],$$

или, учитывая (9),

$$y' = f(x)^{h(x)}h'(x)\ln f(x) + h(x)f(x)^{h(x)-1}f'(x). \qquad (4.9.11)$$

Заметим, что в формуле (11) справа стоит сумма двух членов: первый член

$f(x)^{h(x)} h'(x) \ln f(x)$ есть производная выражения f^h, вычисленная в предположении, что переменной является лишь h, а f — постоянная; второй член $h(x) f(x)^{h(x)-1} f'(x)$ — это производная выражения f^h, вычисленная в предположении, что f — переменная, а h — постоянная. Здесь подтверждается общий принцип, высказанный в § 2.

Заметим, наконец, что в равенстве $y = \log_a x$ (как и в равенстве $y = a^x$) величины x, y и a обязательно безразмерны. Если x — размерная величина (например, длина), то число x задается лишь с точностью до постоянного множителя, зависящего от выбора единицы измерения. Соответственно этому величина $\log_a x$ (при любом фиксированном основании логарифмов a) определена лишь с точностью до постоянного слагаемого (до *аддитивной постоянной*, как говорят в подобных случаях математики); в этом отношении она подобна неопределенному интегралу [3]. Если же обе величины x и y — размерные, то правильно будет писать $y = l_2 \log_a \frac{x}{l_1}$, где множители l_1 и l_2 характеризуют единицы измерения для x и y: ведь взять логарифм можно лишь от безразмерной величины, скажем от числа 100, получаемого как отношение (расстояние AB)/(единица измерения l_1) (например, метр). Значение логарифмической функции также безразмерно; поэтому размерная величина y обязательно равна $l_2 \log_a \left(\frac{x}{l_1}\right)$, т. е. $\log_a \frac{x}{l_1}$ единиц l_2. Эти достаточно простые соображения полезно иметь в виду, ибо логарифмическая (как и экспоненциальная) зависимость часто встречается в естественнонаучных законах.

Укажем еще, что в силу (3) и (3а) переход к иному основанию системы логарифмов b сводится к умножению всех логарифмов на постоянное число $k (= \log_b a)$, т. е. просто к изменению системы единиц, измеряющих величину $y = \log_a x$. Именно поэтому в большинстве случаев выбор основания системы логарифмов не играет никакой роли, что и позволяет пользоваться именно натуральными логарифмами как самыми простыми.

Упражнения

4.9.1. Чему равен $\log_5 15$? [Указание. Воспользуйтесь формулой (3).]

4.9.2. Выведите формулу для производной: а) произведения двух функций, использовав соотношение $\log(uv) = \log u + \log v$; б) частного двух функций, использовав соотношение $\log \frac{u}{v} = \log u - \log v$.

4.9.3. Найдите производные функций: а) $y = \ln(x+3)$; б) $y = \ln 2x$; в) $y = \ln(x^2 + 1)$; г) $y = \ln\left(x + \frac{1}{x}\right)$; д) $y = \ln(3x^2 - x + 1)$; е) $y = \ln \frac{x-1}{x+1}$; ж) $y = \ln \frac{\sqrt{x}}{\sqrt[3]{x+1}}$; з) $y = x \ln x$; и) $y = x^3 \ln(x+1)$; к) $y = x^x$; л) $y = \sqrt{x}^{\sqrt{x^2-1}}$.

§ 10. Тригонометрические функции

В этом параграфе мы найдем производные тригонометрических функций.

Тригонометрические функции определяются как отношения отрезков — отношения сторон прямоугольного треугольника или отношения определенных отрезков в круге («линия синуса», «линия конуса» и т. д.) к радиусу круга. Поэтому тригонометрические функции безразмерны; их аргументом служит также безразмерная величина — угол. Естественной мерой угла, употребляемой в высшей математике, является *радиан* — центральный угол в 1 радиан стягивает в круге дугу, равную своему радиусу, а центральному углу в t радиан отвечает в круге радиуса 1 дуга длины t. О значении тригонометрических функций для описания физических процессов мы скажем во второй части книги (см. гл. 10).

Далее мы всегда будем рассматривать круг с радиусом, равным 1. При этом кратко будем говорить, что в таком круге синус равен длине линии синуса, угол равен длине дуги и т. д. Однако читатель должен помнить, что и тригонометрические функции, и углы безразмерны и не измеряются едини-

[3] По поводу глубокой связи между логарифмической функцией и интегралами (частично объясняющей и эту аналогию) см. формулу (5.2.2) (которую иногда даже кладут в основу определения логарифмов) и обсуждение этой формулы в § 7.6.

цами длины (сантиметрами, дюймами или метрами). Синус равен длине линии синуса (в сантиметрах), деленной на длину радиуса (в сантиметрах); лишь при $r=1$ см он численно равен длине линии синуса. Линии синуса и косинуса показаны на рис. 1.

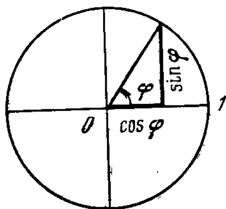
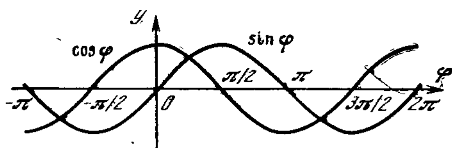


Рис. 4.10.1

Рис. 4.10.2



Напомним вид графиков синуса и косинуса в зависимости от угла (рис. 2). Период синуса, так же как и период косинуса, равен $2\pi\approx 6{,}28$: ведь изменение угла на 2π означает полный оборот радиуса окружности.

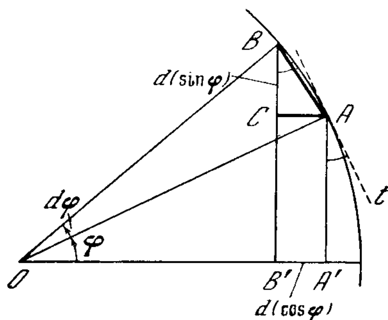


Рис. 4.10.3

Найдем производные синуса и косинуса геометрически. На рис. 3 конец радиуса, отвечающего углу φ, обозначен A, конец радиуса, отвечающего близкому углу $\varphi+d\varphi$, обозначен B. Таким образом, длина дуги AB равна $d\varphi$. Опустим из A перпендикуляр AC на линию синуса BB' угла $\varphi+d\varphi$. Как видно из рис. 3,

$$AA'=\sin\varphi,\quad BB'=\sin(\varphi+d\varphi),$$
$$BC=\sin(\varphi+d\varphi)-\sin\varphi=d(\sin\varphi)$$

(и здесь, конечно, последнее равенство является «точным в пределе» при $d\varphi\to 0$). Далее

$$OA'=\cos\varphi,\quad OB'=\cos(\varphi+d\varphi),$$
$$A'B'=AC=\cos\varphi-\cos(\varphi+d\varphi)=$$
$$=-d(\cos\varphi).$$

Так как угол $d\varphi$ мал, то длина $d\varphi$ дуги AB практически не отличается от длины хорды AB; можно также считать, что угол ABC, образованный хордой AB и вертикалью BCB', равен φ[1]. Из рассмотрения треугольника ABC найдем $BC=AB\cos\varphi$, $AC=AB\sin\varphi$. Таким образом,

$$d(\sin\varphi)=\cos\varphi\, d\varphi,\quad -d(\cos\varphi)=\sin\varphi\, d\varphi$$

и, следовательно,

$$\frac{d(\sin\varphi)}{d\varphi}=\cos\varphi,\quad \frac{d(\cos\varphi)}{d\varphi}=-\sin\varphi. \tag{4.10.1}$$

Такой простой и наглядный способ вычисления производной синуса и косинуса вполне убеждает начинающего читателя. С другой стороны, опытный и все знающий читатель (на которого эта книга не рассчитана) отметит те трудности, которые встречались по ходу вывода. При малом, но конечном (не бесконечно малом!) угле $\Delta\varphi$ длина хорды AB меньше длины дуги: можно показать, что

$$AB=\Delta\varphi-\frac{(\Delta\varphi)^3}{12}+\ldots$$

Далее, угол CBA на самом деле равен $\varphi+\frac{\Delta\varphi}{2}$, а не просто φ, следовательно,

$$BC=AB\cos\left(\varphi+\frac{\Delta\varphi}{2}\right)=$$
$$=\left(\Delta\varphi-\frac{(\Delta\varphi)^3}{12}+\ldots\right)\left(\cos\varphi-\frac{\Delta\varphi}{2}\sin\varphi+\ldots\right).$$

Поэтому выписанные выше равенства, строго говоря, не точны. Однако они верны «с точностью до малых первого порядка» по отношению к малой величине $\Delta\varphi$, т. е. при условии пренебрежения членами второго и высшего порядков относительно $\Delta\varphi$. Поэтому если для приращений верны лишь приближенные равенства $\Delta(\sin\varphi)\approx\cos\varphi\cdot\Delta\varphi$ и $\Delta(\cos\varphi)\approx\approx-\sin\varphi\cdot\Delta\varphi$, то соответствующие равенства для дифференциалов уже точны, и, значит, равенства (1) для производных также являются точными.

Описанный вывод формул для производных тригонометрических функций равносилен сле-

[1] Касательная t к окружности в точке A образует с вертикалью AA' угол $A'At$, равный φ ($\angle A'At=\angle AOA'$ как углы со взаимно перпендикулярными сторонами!); угол же ABC практически не отличается от указанного, ибо хорда AB по направлению очень близка к t.

дующим соображениям, опирающимся на использование механических представлений (и, в частности, апеллирующим к понятию *вектора*). Представим себе точку $M=M(t)$, движение которой по плоскости описывается параметрическими уравнениями:

$$x=\varphi(t), \quad y=\psi(t), \tag{4.10.2}$$

где t — это время (см. § 1.8). В таком случае скорости изменения координат x и y точки M будут задаваться проивводными $x'=\frac{dx}{dt}$ и $y'=\frac{dy}{dt}$. При этом вектор $\frac{\mathbf{MM_1}}{\Delta t}=\mathbf{v}_{\text{ср}}=\left(\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t}\right)$ $\left(\text{где } \frac{\Delta x}{\Delta t} \text{ и } \frac{\Delta y}{\Delta t} \text{ — координаты вектора, рис. 4}\right)$ будет характеризовать среднюю скорость на отрезке MM_1 пути (здесь $M_1=M(t+\Delta t)$); вектор же $\mathbf{v}=\left(\frac{dx}{dt}, \frac{dy}{dt}\right)$ задает мгновенную скорость в точке M пути: он направлен, очевидно, по касательной к траектории движения в точке M, и его длина равна $\lim\limits_{\Delta t\to 0}\frac{\Delta s}{\Delta t}=\frac{ds}{dt}$, где $\Delta s\approx|MM_1|$ — приращение пути s за время Δt.

Пусть теперь траекторией движения точки M служит окружность радиуса 1, по которой точка движется равномерно, пробегая за единицу времени единицу пути, скажем 1 см за 1 с. Тогда формулы (2) примут вид:

$$x=\cos t, \quad y=\sin t \tag{4.10.3}$$

и скорость $\mathbf{v}$ движения будет задаваться вектором с координатами $\frac{d(\cos t)}{dt}$, $\frac{d(\sin t)}{dt}$. Но ясно, что вектор скорости $\mathbf{v}$ будет иметь здесь единичную длину (ибо приращение Δs пути по условию равно приращению Δt времени) и будет направлен по касательной к окружности (см. рис. 4). Координаты этого вектора $\mathbf{v}$, перпендикулярного радиусу $\mathbf{OM}=\mathbf{r}$ окружности, будут, очевидно, равны $-\sin t$ и $\cos t$ (ср. прямоугольные треугольники AOM и VMP со взаимно перпендикулярными сторонами на рис. 4: они имеют одинаковые гипотенузы $OM=VM=1$ и равные острые углы $\widehat{AOM}=\widehat{MVP}=t$; знак минус у координаты $-\sin t$ вектора $\mathbf{v}$ указывает на направление отрезка $\mathbf{MP}$). Таким образом, мы снова приходим к тем же формулам (1):

$$\frac{d(\cos t)}{dt}=-\sin t, \quad \frac{d(\sin t)}{dt}=\cos t.$$

Наконец, укажем еще совсем иной (и более формальный) способ вычисления производных функций $\sin\varphi$ и $\cos\varphi$, не требующий обращения к чертежу. Согласно общим формулам приращение синуса $\Delta\sin\varphi=\sin(\varphi+\Delta\varphi)-\sin\varphi$. Вспомним формулу синуса суммы двух углов

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta$$

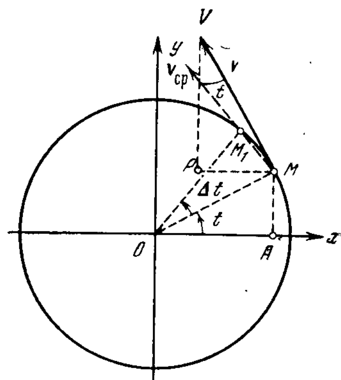


Рис. 4.10.4

и применим ее к $\sin(\varphi+\Delta\varphi)$. Получим

$$\sin(\varphi+\Delta\varphi)=\sin\varphi\cos\Delta\varphi+\cos\varphi\sin\Delta\varphi,$$

откуда

$$\Delta\sin\varphi=\sin\varphi\cos\Delta\varphi+\cos\varphi\sin\Delta\varphi-\sin\varphi=\cos\varphi\sin\Delta\varphi-\sin\varphi(1-\cos\Delta\varphi).$$

Составим отношение приращений:

$$\frac{\Delta\sin\varphi}{\Delta\varphi}=\cos\varphi\frac{\sin\Delta\varphi}{\Delta\varphi}-\sin\varphi\frac{1-\cos\Delta\varphi}{\Delta\varphi}. \tag{4.10.4}$$

Теперь нужно перейти к пределу при $\Delta\varphi\to 0$. Известно, что при очень малых углах α синус практически равен дуге: $\sin\alpha\approx\alpha$, т. е. $\sin\Delta\varphi\approx\Delta\varphi$ при $\Delta\varphi\to 0$. Другими словами,

$$\lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi}=1.$$

Второй член в выражении (4) надо сперва преобразовать по известной формуле $1-\cos 2\alpha=2\sin^2\alpha$, поэтому $1-\cos\Delta\varphi=2\sin^2\left(\frac{\Delta\varphi}{2}\right)$. В этой формуле при малом $\Delta\varphi$ имеем $\sin\left(\frac{\Delta\varphi}{2}\right)\approx\frac{\Delta\varphi}{2}$. Поэтому

$$\frac{1-\cos\Delta\varphi}{\Delta\varphi}\approx\frac{2\left(\frac{\Delta\varphi}{2}\right)^2}{\Delta\varphi}=\frac{\Delta\varphi}{2}.$$

Следовательно, в пределе при $\Delta\varphi\to 0$ второй член пропадает: $\lim\limits_{\Delta\varphi\to 0}\frac{1-\cos\Delta\varphi}{\Delta\varphi}=0$.

Отсюда

$$\lim_{\Delta\varphi\to 0}\frac{\Delta\sin\varphi}{\Delta\varphi}=\frac{d\sin\varphi}{d\varphi}=\cos\varphi.$$

Аналогично

$$\Delta\cos\varphi=\cos(\varphi+\Delta\varphi)-\cos\varphi=$$
$$=\cos\varphi\cos\Delta\varphi-\sin\varphi\sin\Delta\varphi-\cos\varphi=$$
$$=-\cos\varphi(1-\cos\Delta\varphi)-\sin\varphi\sin\Delta\varphi$$

(здесь мы используем формулу для косинуса суммы двух углов) и, значит,

$$\frac{\Delta\cos\varphi}{\Delta\varphi}=-\cos\varphi\frac{1-\cos\Delta\varphi}{\Delta\varphi}-$$
$$-\sin\varphi\frac{\sin\Delta\varphi}{\Delta\varphi}.$$

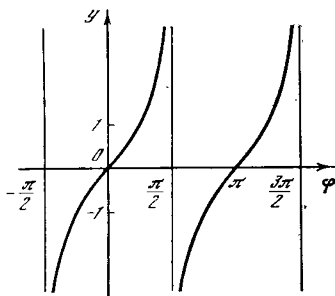


Рис. 4.10.5

А так как мы уже знаем, что

$$\lim_{\Delta\varphi\to 0}\frac{1-\cos\Delta\varphi}{\Delta\varphi}=0,\ \lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi}=1,$$

то окончательно получаем

$$\frac{d\cos\varphi}{d\varphi}=\lim_{\Delta\varphi\to 0}\frac{\Delta\cos\varphi}{\Delta\varphi}=$$
$$=-\cos\varphi\cdot 0-\sin\varphi\cdot 1=-\sin\varphi.$$

Соотношения (1) справедливы для любых углов, а не только для углов в первой четверти. Полезно также, глядя на графики функций $\sin x$ и $\cos x$, проверить, что формулы (1) указывают правильный знак производных при любых x.

Проверим еще формулы (1) при малых углах. При малом φ геометрически очевидно, что $\sin\varphi\approx\varphi$, $\cos\varphi\approx 1$, где приближенное равенство $\approx$, так же как и во многих других случаях, означает «равно с точностью до малых первого порядка». Первая из формул (1) при малом φ дает $\frac{d(\sin\varphi)}{d\varphi}\approx 1$, а вторая дает $\frac{d(\cos\varphi)}{d\varphi}\approx-\varphi$. Поэтому $\frac{d\cos\varphi}{d\varphi}=0$ при $\varphi=0$; равенство нулю производной соответствует тому, что косинус имеет максимум при $\varphi=0$. Равенство же $\frac{d(\sin\varphi)}{d\varphi}=1$ при $\varphi=0$ указывает, что при малых φ величина $\sin\varphi$ растет примерно как φ.

Зная производные функций $y=\sin x$ и $y=\cos x$ и используя соотношения, связывающие тригонометрические функции, легко найти производные всех остальных тригонометрических функций. Так, например, $\operatorname{tg}x=\frac{\sin x}{\cos x}$. Поэтому по формуле дифференцирования дроби

$$\frac{d\operatorname{tg}x}{dx}=\frac{d(\sin x/\cos x)}{dx}=$$
$$=\frac{\cos x\cos x-\sin x(-\sin x)}{\cos^2 x}.$$

Отсюда

$$\frac{d\operatorname{tg}x}{dx}=\frac{\cos^2 x+\sin^2 x}{\cos^2 x}=\frac{1}{\cos^2 x}. \tag{4.10.5}$$

Из рис. 5, на котором изображен график $\operatorname{tg}x$, видно, что функция $y=\operatorname{tg}x$ при любых x имеет положительную производную. Вблизи точек разрыва $\left(x=\frac{\pi}{2},\ x=\frac{3\pi}{2},\ \ldots\right)$ производная неограниченно возрастает — и здесь «уходит в бесконечность» как сама функция, так и ее производная. Оба эти вывода вполне согласуются с формулой (5).

Пользуясь совершенно аналогичным приемом, находим

$$\frac{d(\operatorname{ctg}x)}{dx}=-\frac{1}{\sin^2 x}. \tag{4.10.6}$$

Производные тангенса и котангенса можно найти и непосредственно. Заметим, что

$$\operatorname{tg}\alpha-\operatorname{tg}\beta=\frac{\sin\alpha}{\cos\alpha}-\frac{\sin\beta}{\cos\beta}=$$
$$=\frac{\sin\alpha\cos\beta-\sin\beta\cos\alpha}{\cos\alpha\cos\beta}=\frac{\sin(\alpha-\beta)}{\cos\alpha\cos\beta}.$$

Отсюда

$$\Delta\operatorname{tg}\varphi=\operatorname{tg}(\varphi+\Delta\varphi)-\operatorname{tg}\varphi=\frac{\sin\Delta\varphi}{\cos(\varphi+\Delta\varphi)\cos\varphi}. \tag{4.10.7}$$

Имея в виду, что $\lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi}=1$, получим из (7)

$$\frac{d(\operatorname{tg}\varphi)}{d\varphi}=\lim_{\Delta\varphi\to 0}\frac{\Delta\operatorname{tg}\varphi}{\Delta\varphi}=$$
$$=\lim_{\Delta\varphi\to 0}\frac{\sin\Delta\varphi}{\Delta\varphi}\frac{1}{\cos(\varphi+\Delta\varphi)\cos\varphi}=$$
$$=1\cdot\frac{1}{\cos^2\varphi}=\frac{1}{\cos^2\varphi}.$$

Упражнения

Найдите производные функций:

4.10.1. $y=\sin(2x+3)$. **4.10.2.** $y=\cos(x-1)$. **4.10.3.** $y=\cos(x^2-x+1)$. **4.10.4.** $y=\sin^2 x$. **4.10.5.** $y=\sin 3x\cos^2 x$. **4.10.6.** $y=(\sin 2x)^x$. **4.10.7.** $y=x\operatorname{tg} x$. **4.10.8.** $y=e^{\operatorname{tg} 2x}$. **4.10.9.** $y=\operatorname{ctg}\frac{x}{2}$.

§ 11. Обратные тригонометрические функции

Новые очень интересные результаты получаются при рассмотрении *обратных тригонометрических функций.* Напомним читателю определения этих функций. Функция

$$y=\operatorname{Arc}\sin x \tag{4.11.1}$$

представляет собой угол y, синус которого равен x:

$$\sin y=x. \tag{4.11.2}$$

Равенство (1) означает в точности то же самое, что и (2). Аналогично функция

$$y=\operatorname{Arc}\operatorname{tg} x \tag{4.11.3}$$

представляет собой угол y, тангенс которого равен x:

$$\operatorname{tg} y=x. \tag{4.11.4}$$

Совершенно так же определяются функции $y=\operatorname{Arc}\cos x$ (условием $x=\cos y$) и $y=\operatorname{Arc}\operatorname{ctg} x$ (условием $x=\operatorname{ctg} y$). Заметим, что функции $y=\operatorname{Arc}\sin x$ и $y=\operatorname{Arc}\cos x$ имеют смысл только при $-1\leqslant x\leqslant 1$ (ср. (2)). Функции $y=\operatorname{Arc}\operatorname{tg} x$ и $y=\operatorname{Arc}\operatorname{ctg} x$ имеют смысл при всех значениях x.

Рассмотрим подробнее функцию $y=\operatorname{Arc}\sin x$. Пусть, например, $x=\frac{1}{2}$, $y=\operatorname{Arc}\sin\frac{1}{2}$. Мы можем считать, что $y=\frac{\pi}{6}$, так как $\sin\frac{\pi}{6}=\frac{1}{2}$; однако можно взять и $y=\frac{5\pi}{6}$, так как $\sin\frac{5\pi}{6}$ тоже равен $\frac{1}{2}$. Возможно считать $y=\frac{13\pi}{6}$, $y=\frac{17\pi}{6}$ и т. д. Мы видим, что одному значению x отвечает бесчисленное множество значений y: функция $\operatorname{Arc}\sin x$ является *многозначной* (ср. с. 39) и даже *бесконечнозначной*. Все эти свойства функции $y=\operatorname{Arc}\sin x$ видны из ее графика (рис. 1).

Выделим участок графика функции $x=\sin y$, на котором эта функция монотонна; обычно в качестве такого участка выбирают отрезок кривой $y=\operatorname{Arc}\sin x$, для которого $-\frac{\pi}{2}\leqslant y\leqslant\frac{\pi}{2}$. Эта часть кривой отвечает так называемым *главным значениям* функции $\operatorname{Arc}\sin x$, которые обозначаются $y=\arcsin x$. Если ограничиться рассмотрением функции $y=\arcsin x$, то каждому x соответствует только одно

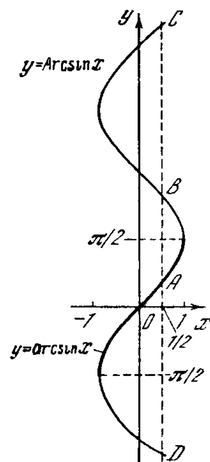


Рис. 4.11.1

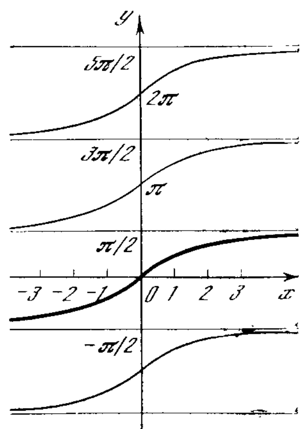


Рис. 4.11.2

значение y; таким образом, функция $\arcsin x$ уже *однозначна.*

Главное значение арктангенса определяется аналогично:

$$-\frac{\pi}{2}<\operatorname{arc}\operatorname{tg} x<\frac{\pi}{2}$$

(ибо функция $x=\operatorname{tg} y$ монотонна на участке $-\frac{\pi}{2}<y<\frac{\pi}{2}$, рис. 2).

Найдем производную функции $y=\arcsin x$. Воспользуемся тем, что арксинус есть функция, обратная синусу: $y=\arcsin x$, $x=\sin y$,

$$x'(y)=\frac{dx}{dy}=\cos y, \tag{4.11.5}$$

$$y'(x)=\frac{dy}{dx}=\frac{1}{x'(y)}=\frac{1}{\cos y}.$$

Однако аргументом здесь мы считаем не y, а x, поэтому $\frac{dy}{dx}$ следует выразить через x, а не через y, как в (5).

Из соотношения $\sin^2 y+\cos^2 y=1$ вытекает, что $\cos y=\pm\sqrt{1-\sin^2 y}$. Так как мы рассматриваем главное значение

арксинуса, то $-\frac{\pi}{2} \leqslant y \leqslant \frac{\pi}{2}$ и, значит, $\cos y \geqslant 0$; поэтому перед корнем берем знак плюс: $\cos y = \sqrt{1-\sin^2 y}$. Так как $\sin y = x$ согласно (2), то $\cos y = \sqrt{1-x^2}$. Подставляя это значение в (5), находим

$$\frac{dy}{dx} = \frac{1}{\sqrt{1-x^2}}, \quad \text{или}$$

$$\frac{d(\arcsin x)}{dx} = \frac{1}{\sqrt{1-x^2}}. \tag{4.11.6}$$

Формулой (6) можно пользоваться только для главного значения функции; для других участков кривой надо внимательно следить за знаком выражения $\sqrt{1-x^2}$ (его ведь тоже можно считать двузначным). Действительно, при одном и том же значении x на различных участках кривой производная имеет разные знаки: так, в точках A и C графика функции $y = \operatorname{Arc}\sin x$ (см. рис. 1) производная положительна, а в точках B и D — отрицательна.

Найдем теперь производную $\frac{d(\operatorname{arc\,tg} x)}{dx}$.

Если $y = \operatorname{arc\,tg} x$, то $x = \operatorname{tg} y$. Отсюда аналогично предыдущему находим:

$$\begin{gathered} x'(y) = \frac{dx}{dy} = \frac{1}{\cos^2 y}, \\ y'(x) = \frac{dy}{dx} = \frac{1}{x'(y)} = \cos^2 y. \end{gathered} \tag{4.11.7}$$

Но из тригонометрии известно, что

$$\operatorname{tg}^2 y + 1 = \frac{1}{\cos^2 y},$$

поэтому

$$\frac{1}{\cos^2 y} = 1 + x^2.$$

Пользуясь (7), получаем окончательно

$$\frac{dy}{dx} = \frac{d(\operatorname{arc\,tg} x)}{dx} = \frac{1}{1+x^2}. \tag{4.11.8}$$

Для любой другой ветви арктангенса (см. рис. 2) формула (8) остается справедливой, так как любая другая ветвь получается из основной параллельным переносом, что не меняет величины производной.

Упражнения

Найдите производные функций:

4.11.1. $y = \arccos x$ (эта функция определяется условием $0 \leqslant y \leqslant \pi$). **4.11.2.** $y = \operatorname{arc\,ctg} x$ (здесь $0 < y < \pi$). **4.11.3.** $y = \arcsin 2x$. **4.11.4.** $y = \operatorname{arc\,tg}(3x+1)$. **4.11.5.** $y = \operatorname{arc\,tg}(x^2 - x)$. **4.11.6.** $y = e^{\operatorname{arc\,tg}\sqrt{x}}$.

§ 12. Производная функции, заданной неявно

Неявное задание функции $y(x)$ — это задание ее выражением вида

$$F(x, y) = 0. \tag{4.12.1}$$

Если соответствующее уравнение можно разрешить относительно x или y, то мы снова придем к обычному (явному) заданию функции. Однако иногда такое решение приводит к сложным формулам, а в других случаях его и вовсе нельзя найти. Так, например, уравнение окружности в форме

$$x^2 + y^2 - 1 = 0 \tag{4.12.2}$$

проще, чем следующее из него явное выражение для функции $y(x)$:

$$y = \pm\sqrt{1-x^2}. \tag{4.12.3}$$

Если в (1) левая часть — произвольный многочлен, то в общем случае это уравнение нельзя разрешить относительно x или относительно y. В качестве другого примера можно привести простое с виду уравнение

$$F(x, y) = x\sin x + y\sin y - \pi = 0. \tag{4.12.4}$$

Однако и в тех случаях, когда мы не можем выписать формулу, дающую прямой способ вычисления y по данному x, все равно y можно считать функцией x: при каждом x можно, решая уравнение численно, найти соответствующее y, можно построить кривую (1) в плоскости x, y. Возможно, что кривая будет существовать не при всех x (в случае окружности (2), например, лишь при $-1 < x < 1$); функция может оказаться многозначной, т. е. данному x могут отвечать более одного значения y (в случае окружности, например, два значения, в соответствии со знаком $\pm$ у корня квадратного). Однако эти осложнения не отменяют основного факта: уравнение $F(x, y) = 0$ определяет y как функцию x.

Как найти производную $\frac{dy}{dx}$? Можно ли это сделать, не решив уравнение (1), т. е. не выразив $y(x)$ явно?

Процедуру отыскания величины y' по равенству (1) знал еще Ньютон. Пусть x, y удовлетворяют уравнению (1). Подставим в это уравнение «соседние»

значения $x+\Delta x$, $y+\Delta y$, которые также должны ему удовлетворять:

$$F(x+\Delta x,\ y+\Delta x)=0. \qquad (4.12.5)$$

Запишем, пользуясь (1),

$$F(x+\Delta x,\ y+\Delta y)=[F(x+\Delta x,\ y+\Delta y)-$$
$$-F(x+\Delta x,\ y)]+[F(x+\Delta x,\ y)-$$
$$-F(x,\ y)]. \qquad (4.12.6)$$

Разность $F(x+\Delta x,y)-F(x,y)$ представляет собой приращение функции $F(x,y)$, рассматриваемой как функция одной переменной x при неизменном y. Это приращение, как мы знаем, в пределе[1] может быть выражено так (см. формулу (1.2)):

$$F(x+\Delta x,\ y)-F(x,\ y)\approx$$
$$\approx\frac{dF(x,\ y)}{dx}\bigg|_{y=\text{const}}\Delta x.$$

При вычислении производной $\frac{dF(x,\ y)}{dx}\Big|_{y=\text{const}}$ функции $F(x,\ y)$ по x аргумент y этой функции мы закрепляем, считаем постоянным; это и означает запись $y=\text{const}$ после обозначения $\frac{dF(x,\ y)}{dx}$ производной. Поскольку такая запись громоздка, вместо нее употребляют более короткую: $\frac{\partial F(x,\ y)}{\partial x}$; здесь круглые буквы ∂ (вместо обычного d) как раз и символизируют тот факт, что производная ищется при постоянном y, т. е. в условиях, когда $F(x,\ y)$ рассматривается как функция одной переменной x (зависящая, правда, от параметра y). Производная

$$\frac{\partial F(x,\ y)}{\partial x}=\frac{dF(x,\ y)}{dx}\bigg|_{y=\text{const}}=$$
$$=\lim_{\Delta x\to 0}\frac{F(x+\Delta x,\ y)-F(x,\ y)}{\Delta x}$$

называется *частной производной* функции $F(x,\ y)$ по переменной (аргументу функции) x; аналогично

$$\frac{\partial F(x,\ y)}{\partial y}=\frac{dF(x,\ y)}{dy}\bigg|_{x=\text{const}}=$$
$$=\lim_{\Delta y\to 0}\frac{F(x,\ y+\Delta y)-F(x,\ y)}{\Delta y}$$

— частная производная от $F(x,\ y)$ по y.

Итак,

$$F(x+\Delta x,\ y)-F(x,\ y)\approx\frac{\partial F(x,\ y)}{\partial x}\Delta x.$$

Аналогично для первой разности в (6) можно написать

$$F(x+\Delta x,\ y+\Delta y)-F(x+\Delta x,\ y)\approx$$
$$\approx\frac{\partial F(x+\Delta x,\ y)}{\partial y}\Delta y.$$

Условие (5) дает

$$\frac{\partial F(x,\ y)}{\partial x}\Delta x+\frac{\partial F(x+\Delta x,\ y)}{\partial y}\Delta y\approx 0,$$

или

$$\frac{\Delta y}{\Delta x}\approx-\frac{\dfrac{\partial F(x,\ y)}{\partial x}}{\dfrac{\partial F(x+\Delta x,\ y)}{\partial y}}.$$

Переходя к пределу при $\Delta x\to 0$, слева получим производную, а справа при этом можно будет отбросить Δx. Окончательно

$$\frac{dy}{dx}=-\frac{\dfrac{\partial F(x,\ y)}{\partial x}}{\dfrac{\partial F(x,\ y)}{\partial y}}. \qquad (4.12.7)$$

Обратите внимание на знак минус в формуле (7) и на то, что в данном случае просто «сократить» $\partial F(x,\ y)$ в числителе и знаменателе нельзя.

Покажем применение (7) на примере уравнения (2). Имеем $F(x,\ y)=x^2+y^2-1$; тогда

$$\frac{\partial F(x,\ y)}{\partial x}=2x,\quad \frac{\partial F(x,\ y)}{\partial y}=2y, \qquad (4.12.8)$$

$$\frac{dy}{dx}=-\frac{2x}{2y}=-\frac{x}{y}.$$

Легко убедиться, что этот результат совпадает с тем, что получится, если вычислить производную функции (3).

Найдем производную в случае (4):

$$\frac{\partial F(x,\ y)}{\partial x}=\sin x+x\cos x,$$

$$\frac{\partial F(x,\ y)}{\partial y}=\sin y+y\cos y,$$

$$\frac{dy}{dx}=-\frac{\sin x+x\cos x}{\sin y+y\cos y}.$$

Таким образом, в выражение производной неявной функции входят обе величины x и y. Чтобы найти ее численно, нужно найти значение y, отвечающее заданному x. Но если бы мы не имели формулы (7), то для нахождения производной нам пришлось бы

[1] Смысл выражения «равно в пределе» (т. е. при малом Δx или Δy) разъяснен подробно в § 2.4.

находить численно два значения y_2 и y_1, отвечающие двум соседним значениям x_1 и x_2, а затем искать отношение $\frac{y_2-y_1}{x_2-x_1}$ и предел этого отношения[2]. При этом чем ближе x_2 к x_1, тем точнее пришлось бы вычислять y_2 и y_1, а это часто вовсе не так просто. Использование же формулы (7) обычно не представляет труда.

Если $F(x, y)=0$ приводит к неоднозначной функции $y(x)$, т. е. если одному значению x отвечают два или больше значений y (несколько ветвей кривой), то выражение (7) при подстановке в него данного x и разных y дает значения производной в соответствующих точках линии $F(x, y)=0$. Читателю предлагается проверить это на примере уравнения окружности (2), для которого производная выражается формулой (8).

Для нахождения производной функции, заданной неявно, нам понадобилось понятие **ч а с т н о й** производной. Оно имеет большое значение, так как необходимо для изучения функций нескольких переменных (которых, впрочем, мы в этой книге почти не касаемся). Неявно мы уже пользовались понятием частной производной: так, определенный интеграл $I=\int_a^b f(x)\,dx=I(a, b)$ есть функция двух переменных; выше мы находили частные производные $\frac{\partial I}{\partial a}\,(=-f(a))$ и $\frac{\partial I}{\partial b}\,(=f(b))$ по одному из пределов, считая второй предел интегрирования закрепленным (постоянным). Даже в таких элементарных вопросах, как производная произведения нескольких функций $y=h(x)g(x)$ или, например, производная степени $y=h(x)^{g(x)}$ (см. § 9), по существу, у нас «работали» частные производные. В самом деле, когда мы говорили, что y складывается из членов, получающихся при взятии производной по переменной x, входящей в $h(x)$, и производной по x, входящей в выражение $g(x)$, то мы имели в виду частные производные. Соответствующее общее правило можно записать так: если

$$y=F[g(x), h(x)],$$

то

$$y'=\frac{dy}{dx}=\frac{\partial F}{\partial g}\frac{dg}{dx}+\frac{\partial F}{\partial h}\frac{dh}{dx}.$$

Упражнения

4.12.1. Найдите производную $\frac{dy}{dx}$ функции, заданной уравнением (4), в точках: а) $x=\frac{\pi}{2}$, $y=\frac{\pi}{2}$; б) $x=-\frac{\pi}{2}$, $y=\frac{\pi}{2}$.

4.12.2. Найдите производную $\frac{dy}{dx}$ функции, заданной уравнением $x^3+3x+y^3+3y-8=0$, в точке $x=y=1$.

---

[2] Дополнительные трудности возникают, когда уравнение (1) задает м н о г о з н а ч н у ю функцию $y(x)$, ибо тут необходимо добиться «согласованности» значений $y_1(x_1)$ и $y_2(x_2)$ — иначе при малой разности x_2-x_1 разность y_2-y_1 может оказаться большой.

Глава 5

ТЕХНИКА ИНТЕГРИРОВАНИЯ[1]

§ 1. Постановка задачи

В гл. 3 мы познакомились с понятием интеграла. При этом была выяснена тесная связь между двумя различными на первый взгляд задачами:

1) нахождение суммы большого числа малых слагаемых вида $v(t)\Delta t$ (или $v(t)dt$);

2) нахождение функции $z(t)$, производная $v(t)$ которой нам известна:

$$\frac{dz}{dt} = v(t). \qquad (5.1.1)$$

Задачи физики, техники, химии, да и математики возникают как задачи типа 1), т. е. задачи суммирования большого числа малых величин. Такая постановка вопроса более наглядна; сама его формулировка уже подсказывает простой, хотя и приближенный, путь вычисления интересующей нас величины — с помощью прямого суммирования тех (малых) слагаемых, о которых в задаче идет речь. Однако этот «прямой» метод решения задач 1) не позволяет выразить ответ в виде формулы, и высшая математика возникла, когда была установлена связь между задачами 1) и 2), что открыло путь к общим приемам (алгоритмам) решения задач 1).

Постановка задач типа 2) является более искусственной. Однако у нее есть свои преимущества. Нахождение производных оказалось простым делом, сводящимся к четырем-пяти формулам (производные функций x^n, e^x, $\ln x$, $\sin x$, $\cos x$ и т. д.) и двум-трем общим правилам. Поэтому легко можно найти производные большого числа функций. Каждый раз, когда найдена производная какой-либо функции $\frac{dz}{dt} = v$, можно зарегистрировать, что для функции v известен интеграл z (см. § 2). Таким образом, набирается много отдельных частных случаев, для которых можно решить задачу типа 2). Кроме того, с помощью тождественных алгебраических преобразований удается для нескольких простых типов функций v (например, когда v есть сумма функций, интегралы которых известны) найти общие правила решения задачи 2) (см. § 3).

Это невозможно, однако, сделать для всех элементарных функций, так что интегрировать труднее, чем дифференцировать (находить производные). Тем не менее формулы, полученные для некоторых интегралов при второй постановке задачи, очень важны. Если уже удалось для данной $v(t)$ найти интеграл (неопределенный интеграл, или первообразную функцию) $z(t)$, то тогда все задачи в первой постановке, все суммы, т. е. все определенные интегралы $\int\limits_a^b v(t)\,dt$, оказываются выраженными простыми формулами, включающими функцию z:

$$\int\limits_a^b v(t)\,dt = z(b) - z(a).$$

Такой результат является гораздо более полным, более точным и ценным, чем итоги каждого отдельного численного расчета суммы, т. е. определенного интеграла $\int\limits_a^b v(t)\,dt$ в каких-то пределах от a до b. Поэтому нашей целью в первую очередь будет именно решение задачи 2), задачи интегрирования во второй ее постановке.

§ 2. Простейшие интегралы

Выпишем формулы для производных, найденные в предыдущих параграфах, и соответствующие им формулы для (неопределенных) интегралов:

$$\frac{d}{dx}(x^n) = nx^{n-1},$$

$$n\int x^{n-1}dx = x^n + C;$$

[1] Перед чтением этой главы полезно повторить материал гл. 3.

$$\frac{d}{dx}(e^{kx}) = ke^{kx},$$

$$k\int e^{kx}dx = e^{kx} + C;$$

$$\frac{d}{dx}(\ln x) = \frac{1}{x},$$

$$\int \frac{1}{x}dx = \ln x + C;$$

$$\frac{d}{dx}(\sin kx) = k\cos kx,$$

$$k\int \cos kx dx = \sin kx + C;$$

$$\frac{d}{dx}(\cos kx) = -k\sin kx,$$

$$-k\int \sin kx dx = \cos kx + C;$$

$$\frac{d}{dx}(\operatorname{tg} x) = \frac{1}{\cos^2 x},$$

$$\int \frac{1}{\cos^2 x}dx = \operatorname{tg} x + C;$$

$$\frac{d}{dx}(\operatorname{ctg} x) = -\frac{1}{\sin^2 x},$$

$$-\int \frac{1}{\sin^2 x}dx = \operatorname{ctg} x + C;$$

$$\frac{d}{dx}(\arcsin x) = \frac{1}{\sqrt{1-x^2}},$$

$$\int \frac{1}{\sqrt{1-x^2}}dx = \arcsin x + C;$$

$$\frac{d}{dx}(\operatorname{arc\,tg} x) = \frac{1}{1+x^2},$$

$$\int \frac{1}{1+x^2}dx = \operatorname{arc\,tg} x + C.$$

Преобразуем полученные формулы. В первом интеграле обозначим $n-1=m$ (т. е. $n=m+1$) и перепишем его так:

$$\int x^m dx = \frac{1}{m+1}x^{m+1} + C. \qquad (5.2.1)$$

Очевидно, что формула (1) справедлива при всех m, кроме $m=-1$; при $m=-1$ знаменатель обращается в нуль, $x^{m+1}=x^0=1$, так что получается непригодное для расчетов выражение $\frac{1}{0}+C$.

Однако как раз в случае $m=-1$, т. е. для интеграла $\int\frac{1}{x}dx$, имеет место формула[1]

$$\int \frac{1}{x}dx = \ln x + C. \qquad (5.2.2)$$

Эта формула справедлива лишь для положительных значений x, так как $\ln x$ имеет смысл лишь при $x>0$. При $x<0$ величина $\ln x$ не имеет смысла, но имеет смысл $\ln(-x)$. Так как

$$\frac{d\ln(-x)}{dx} = \frac{1}{-x}\cdot(-1) = \frac{1}{x},$$

то при $x<0$ имеем $\int\frac{dx}{x} = \ln(-x)+C$.

Обе формулы могут быть объединены в одну:

$$\int \frac{dx}{x} = \ln|x| + C. \qquad (5.2.3)$$

Формула (3) применима на любом промежутке интегрирования, не содержащем $x=0$.

Интеграл от показательной функции запишется так:

$$\int e^{kx}dx = \frac{1}{k}e^{kx} + C. \qquad (5.2.4)$$

Аналогично получаем для синуса и косинуса:

$$\int \sin kx dx = -\frac{1}{k}\cos kx + C, \qquad (5.2.5)$$

$$\int \cos kx dx = \frac{1}{k}\sin kx + C. \qquad (5.2.6)$$

Рассмотрим в заключение пример, показывающий необходимость внимательного рассмотрения функции и опасность формального подхода к задаче интегрирования. Пусть $I=\int_a^b\frac{dx}{x^2}$. Неопределенный интеграл $\int\frac{dx}{x^2} = -\frac{1}{x}+C$; поэтому

$$I = \int_a^b \frac{dx}{x^2} = -\frac{1}{x}\Big|_a^b = -\frac{1}{b}+\frac{1}{a} = \frac{b-a}{ab}. \qquad (5.2.7)$$

Так как подынтегральная функция положительна, результат должен быть положителен, если $b>a$. Ответ действительно положителен при $b>a$, если a и b одного знака. Однако для интеграла $\int_{-1}^{1}\frac{dx}{x^2}$ формула (7) дает явно

[1] В § 7.6 мы специально обсудим неожиданно резкое различие между формулами (1) и (2) для интегралов от степенной функции.

нелепое значение $I=-2$. Причина заключается в том, что подынтегральная функция обращается в бесконечность внутри промежутка интегрирования (при $x=0$) — и в этом же месте терпит бесконечный разрыв функция $-\frac{1}{x}$, являющаяся неопределенным интегралом функции $\frac{1}{x^2}$.

Чтобы разобраться в истинном положении дела, необходимо исключить из всего интервала $-1<x<1$ малую область около «особой точки» $x=0$: $-\varepsilon_1<x<\varepsilon_2$ (ε_1 и ε_2 — малые положительные числа) и рассмотреть

$$K=\int_{-1}^{-\varepsilon_1}\frac{dx}{x^2}+\int_{\varepsilon_2}^{1}\frac{dx}{x^2}.$$

Естественно считать, что наш интеграл I получится из K при стремлении ε_1 и ε_2 к нулю. Но по формуле (7) получаем

$$K=\frac{1-\varepsilon_1}{\varepsilon_1}+\frac{1-\varepsilon_2}{\varepsilon_2}=-2+\frac{1}{\varepsilon_1}+\frac{1}{\varepsilon_2}.$$

Ясно, что при ε_1 и ε_2, стремящихся к нулю, $K\to\infty$.

В других случаях интеграл с подынтегральной функцией, обращающейся в бесконечность на интервале интегрирования, может давать вполне определенный конечный результат. Так, например,

$$\int_0^1\frac{dx}{\sqrt{x}}=2.$$

В самом деле, здесь неопределенный интеграл равен

$$\int\frac{dx}{\sqrt{x}}=\int x^{-1/2}dx=\frac{1}{1/2}x^{1/2}+$$
$$+C=2\sqrt{x}+C.$$

Поэтому

$$\int_\varepsilon^1\frac{dx}{\sqrt{x}}=2\sqrt{x}\Big|_\varepsilon^1=2-2\sqrt{\varepsilon},$$

а при $\varepsilon\to 0$ это выражение стремится к 2.

Такого рода рассмотрение всегда необходимо при обращении подынтегральной функции в бесконечность; однако здесь мы не будем подробно обсуждать эту важную тему (ср. § 6.3).

Упражнения

Чему равны следующие определенные интегралы:

5.2.1. $\int_0^1 x^5dx$. 5.2.2. $\int_1^2\frac{1}{x^2}dx$.

5.2.3. $\int_0^\pi \sin x dx$. 5.2.4. $\int_0^1\frac{dx}{x^2+1}$.

5.2.5. $\int_{-1}^1 e^x dx$. 5.2.6. $\int_0^{\pi/4}\frac{dx}{\cos^2 x}$.

§ 3. Общие свойства интегралов

Выше, в § 4.1—4.3, были найдены формулы для производной суммы (и разности) функций, производной произведения (и частного) функций, производных сложной и обратной функций. Каждому из этих свойств соответствует определенное свойство, относящееся к интегралам. Следующие три параграфа будут посвящены «интегральным аналогам» рассматриваемых свойств производных.

Для интегралов имеет место равенство

$$\int[Af(x)+Bg(x)]\,dx=$$
$$=A\int f(x)\,dx+B\int g(x)\,dx. \qquad (5.3.1)$$

Для доказательства достаточно продифференцировать стоящее справа выражение: если равенство верно, то мы получим подынтегральную функцию. И в самом деле,

$$\left[A\int f(x)\,dx+B\int g(x)\,dx\right]'=$$
$$=A\left[\int f(x)\,dx\right]'+B\left[\int g(x)\,dx\right]'=$$
$$=Af(x)+Bg(x).$$

Таким образом, равенство (1) доказано. Оно показывает, что *интеграл суммы нескольких слагаемых равен сумме интегралов отдельных слагаемых* и что *постоянные множители можно выносить за знак интеграла.*

В качестве простейшего примера найдем интеграл от м н о г о ч л е н а:

$$\int(a_0x^k+a_1x^{k-1}+a_2x^{k-2}+\ldots+a_{k-2}x^2+$$
$$+a_{k-1}x+a_k)\,dx=a_0\int x^kdx+$$

$$+a_1\int x^{k-1}dx+a_2\int x^{k-2}dx+\ldots$$
$$\ldots+a_{k-2}\int x^2dx+a_{k-1}\int xdx+a_k\int dx=$$
$$=\frac{a_0}{k+1}x^{k+1}+\frac{a_1}{k}x^k+\frac{a_2}{k-1}x^{k-1}+\ldots$$
$$\ldots+\frac{a_{k-2}}{3}x^3+\frac{a_{k-1}}{2}x^2+a_kx+C. \quad (5.3.2)$$

Отсюда, в частности, имеем

$$\int_n^m(a_0x^k+a_1x^{k-1}+a_2x^{k-2}+\ldots+a_{k-2}x^2+$$
$$+a_{k-1}x+a_k)\,dx=\frac{a_0}{k+1}(m^{k+1}-n^{k+1})+$$
$$+\frac{a_1}{k}(m^k-n^k)+\frac{a_2}{k-1}(m^{k-1}-$$
$$-n^{k-1})+\ldots+\frac{a_{k-2}}{3}(m^3-n^3)+$$
$$+\frac{a_{k-1}}{2}(m^2-n^2)+a_k(m-n).$$

Так, например,

$$\int(2x^2-6x+1)\,dx=\frac{2}{3}x^3-3x^2+x+C,$$
$$\int_1^2(2x^2-6x+1)\,dx=\frac{2}{3}(8-1)-$$
$$-3(4-1)+(2-1)=-3\frac{1}{3}.$$

Под знаком интеграла можно делать замену переменной, переходить к новой, более удобной переменной. Подробнее мы остановимся на этом в § 5, однако несколько простейших примеров рассмотрим уже здесь.

1. Найти $\int(ax+b)^n\,dx$, где $n\neq-1$.

Здесь новой переменной удобно объявить стоящую в скобке величину. Назовем ее z:

$$ax+b=z. \quad (5.3.3)$$

При этом надо также от дифференциала dx перейти к дифференциалу dz. Но из (3) следует: $dz=a\,dx$, $dx=\frac{dz}{a}$. Таким образом,

$$\int(ax+b)^n\,dx=\int z^n\frac{dz}{a}=\frac{1}{a}\int z^ndz=$$
$$=\frac{z^{n+1}}{a(n+1)}+C=\frac{(ax+b)^{n+1}}{a(n+1)}+C. \quad (5.3.4)$$

В правильности результата легко убедиться, вычисляя производную правой части:

$$\frac{d}{dx}\left[\frac{(ax+b)^{n+1}}{a(n+1)}+C\right]=\frac{d}{dx}\left[\frac{(ax+b)^{n+1}}{a(n+1)}\right]=$$
$$=\frac{n+1}{a(n+1)}(ax+b)^n\frac{d}{dx}(ax+b)=$$
$$=\frac{(ax+b)^n}{a}a=(ax+b)^n.$$

2. Аналогично выполняется замена переменной и в интеграле $\int\frac{dx}{ax+b}$: $z=ax+b$, $dx=\frac{dz}{a}$, и, следовательно,

$$\int\frac{dx}{ax+b}=\frac{1}{a}\int\frac{dz}{z}=\frac{1}{a}\ln|z|+C=$$
$$=\frac{\ln|ax+b|}{a}+C. \quad (5.3.5)$$

На практике в таких очень несложных примерах преобразования делают менее торжественно, не вводя отдельных обозначений для новых, промежуточных переменных. Вместо (4) пишут, например:

$$\int(ax+b)^n\,dx=\int(ax+b)^n\frac{1}{a}d(ax+b)=$$
$$=\frac{1}{(n+1)a}(ax+b)^{n+1}+C.$$

Приведем теперь несколько более сложный пример интеграла, который удается привести к известным нам интегралам посредством алгебраических преобразований.

Рассмотрим интеграл $\int\frac{dx}{(x-a)(x-b)}$. Заметим, что справедливо тождество

$$\frac{1}{x-a}-\frac{1}{x-b}=\frac{a-b}{(x-a)(x-b)}. \quad (5.3.6)$$

Из (6) следует

$$\frac{1}{(x-a)(x-b)}=\frac{1}{a-b}\left(\frac{1}{x-a}-\frac{1}{x-b}\right).$$

Поэтому

$$\int\frac{dx}{(x-a)(x-b)}=$$
$$=\frac{1}{a-b}\int\left(\frac{1}{x-a}-\frac{1}{x-b}\right)dx=$$
$$=\frac{1}{a-b}[\ln|x-a|-\ln|x-b|]+C=$$
$$=\frac{1}{a-b}\ln\left|\frac{x-a}{x-b}\right|+C. \quad (5.3.7)$$

(формулой (7) можно пользоваться в любом промежутке, не включающем значений $x=a$ и $x=b$). Интеграл от любой алгебраической дроби (от отношения $\frac{P_n(x)}{Q_m(x)}$ многочленов каких-то сте-

пеней n и m) можно выразить в виде комбинации элементарных функций (алгебраических функций, логарифмов и арктангенсов). Некоторые простейшие примеры такого рода имеются в упражнениях.

Упражнения

Найдите (неопределенные) интегралы:

5.3.1. $\int (3x^2 - 2x + 1)\,dx$. 5.3.2. $\int (4x^4 - 3x^3 + x^2 - x)\,dx$. 5.3.3. $\int x(x-1)^2\,dx$.

5.3.4. $\int \frac{x^2+2x+3}{x}\,dx$. 5.3.5. $\int \frac{2x-1}{x-1}\,dx$.

5.3.6. $\int \frac{ax+b}{cx+d}\,dx$. [Указание. Воспользуйтесь тем, что $\frac{ax+b}{cx+d} = \frac{a}{c} + \frac{bc-ad}{c(cx+d)}$.]

5.3.7. $\int \frac{x\,dx}{(x-2)(x-3)}$. [Указание. Воспользуйтесь тождеством $\frac{x}{(x-2)(x-3)} = \frac{A}{x-2} + \frac{B}{x-3}$; числа A и B находятся «методом неопределенных коэффициентов», т. е. приравниванием коэффициентов при одинаковых степенях x после освобождения дробей от знаменателей.] 5.3.8. $\int \frac{dx}{(x-1)(x-2)}$.

5.3.9. $\int \frac{x+1}{x^2-3x+2}\,dx$. 5.3.10. $\int \frac{x+2}{x^3+x}\,dx$. [Указание. Воспользуйтесь тождеством $\frac{x+2}{x^3+x} = \frac{x+2}{x(x^2+1)} = \frac{A}{x} + \frac{B}{x^2+1} + \frac{Cx}{x^2+1}$, где коэффициенты A, B, C, находятся «методом неопределенных коэффициентов» (см. указание к упр. 7); чтобы найти $\int \frac{Cx\,dx}{x^2+1}$, удобно обозначить $x^2+1=z$.]

§ 4. Интегрирование по частям

Пусть $f(x)$ и $g(x)$ — две различные функции переменной x. Формула для производной произведения дает

$$\frac{d}{dx}(fg) = g\frac{df}{dx} + f\frac{dg}{dx}. \tag{5.4.1}$$

Равенство (1) позволяет утверждать, что

$$fg = \int\left(g\frac{df}{dx} + f\frac{dg}{dx}\right)dx = \int f\frac{dg}{dx}\,dx + \int g\frac{df}{dx}\,dx. \tag{5.4.2}$$

В справедливости (2) убеждаемся, находя производную от левой и правой частей этого равенства — при этом мы приходим к верному равенству (1).

Перепишем (2) так:

$$\int f\frac{dg}{dx}\,dx = fg - \int g\frac{df}{dx}\,dx,$$

или, что то же самое,

$$\int f\,dg = fg - \int g\,df. \tag{5.4.3}$$

В чем смысл формулы (3)? К сожалению, не существует правила, выражающего интеграл от произведения двух функций через интегралы от каждого из сомножителей. Однако если для произведения двух функций fw известен интеграл от одного из сомножителей:

$$\int w\,dx = g, \quad w = \frac{dg}{dx}, \tag{5.4.4}$$

то интеграл $\int fw\,dx$ удается выразить через интеграл, в который входит производная $\frac{df}{dx}$. Учитывая (4), перепишем (3) в виде

$$\int fw\,dx = f\left(\int w\,dx\right) - \int\left(\int w\,dx\right)\frac{df}{dx}\,dx. \tag{5.4.5}$$

Так как $\int w\,dx = g$, то последний интеграл в (5) есть $\int g\frac{df}{dx}\,dx$; иногда он бывает проще исходного интеграла $\int fw\,dx$ или сводится к известному интегралу. В частности, если f — степенная функция или многочлен, то $\frac{df}{dx}$ — тоже степенная функция (многочлен), причем степень $\frac{df}{dx}$ на единицу н и ж е степени f.

Формула (3), или (5), называется формулой *интегрирования по частям*. Из нее вытекает и аналогичное соотношение для определенных интегралов:

$$\int_a^b f(x)\,dg(x) = f(x)g(x)\Big|_a^b - \int_a^b g(x)\,df(x) = [f(b)g(b) - f(a)g(a)] - \int_a^b g(x)\,df(x). \tag{5.4.3а}$$

Приведем примеры использования метода интегрирования по частям.

1. Найти $\int xe^x dx$.

Положим $f=x$; тогда

$w=\frac{dg}{dx}=e^x, \quad e^x dx=dg,$

$g=\int e^x dx=e^x, \quad df=dx.$

По формуле (3) имеем

$\int xe^x dx = xe^x - \int e^x dx =$

$= xe^x - e^x + C = e^x(x-1)+C.$

2. Найти $\int x^2 e^x dx$.

Положим $f=x^2$; тогда

$w=\frac{dg}{dx}=e^x, \quad e^x dx=dg,$

$g=\int e^x dx=e^x, \quad df=2xdx.$

Пользуясь формулой (3), получаем

$\int x^2 e^x dx = x^2 e^x - 2\int xe^x dx,$

откуда, учитывая результаты первого примера, имеем

$\int x^2 e^x dx = x^2 e^x - 2xe^x + 2e^x + C =$

$=(x^2-2x+2)e^x + C.$

Для нахождения $\int P_n(x)e^{kx}dx$, где $P_n(x)$ — многочлен степени n, приходится n раз выполнять интегрирование по частям. При этом в ответе получится выражение вида $Q_n(x)e^{kx}$, где $Q_n(x)$ — многочлен степени n. Зная это, можно не выполнять n раз интегрирование по частям, а прямо находить коэффициенты многочлена $Q_n(x)$.

Рассмотрим тот же пример 2: найти $\int x^2e^x dx$. Напишем равенство $\int x^2e^x dx = Q_2(x)e^x + C$ с неизвестными пока коэффициентами квадратичного многочлена Q_2:

$$\int x^2e^x dx = (a_2x^2+a_1x+a_0)e^x + C. \qquad (5.4.6)$$

Составим производную от обеих частей равенства (6):

$x^2e^x = (2a_2x+a_1)e^x + (a_2x^2+a_1x+a_0)e^x,$

или

$x^2e^x = [x^2a_2 + x(2a_2+a_1) + (a_1+a_0)]e^x.$

Приравняем коэффициенты при одинаковых степенях x в многочленах, стоящих справа и слева. Получим: $a_2=1$, $2a_2+a_1=0$, $a_1+a_0=0$, откуда $a_1=-2$, $a_0=2$. Окончательно получим, как и раньше,

$\int x^2e^x dx = (x^2-2x+2)e^x + C.$

Аналогичным приемом можно находить интегралы от функций $P_n(x)\cos kx$ и $P_n(x)\sin kx$, где $P_n(x)$ — многочлен. В обоих случаях ответ имеет вид

$Q_n(x)\cos kx + R_n(x)\sin kx,$

где $Q_n(x)$ и $R_n(x)$ — многочлены, степень которых равна n (или меньше n). Примеры такого рода приведены в упражнениях.

Упражнения

Найдите интегралы:

5.4.1. $\int x\cos x\,dx$. 5.4.2. $\int \ln x\,dx$. 5.4.3. $\int x^2\sin 2x\,dx$. 5.4.4. $\int x^3e^{-x}dx$. 5.4.5. $\int (x^2+x+1)\cos x\,dx$. 5.4.6. $\int (2x^2+1)\cos 3x\,dx$. [Указание. Упр. 6 подробно рассмотрено в «Ответах, указаниях и решениях»; упр. 3—5 аналогичны ему.] 5.4.7. $\int \arcsin x dx$, 5.4.8. $\int \operatorname{arctg} x\,dx$. 5.4.9. $\int e^{2x}\sin 3x\,dx$. 5.4.10. $\int e^x\cos 2x\,dx$.

§ 5. Метод подстановки

Мы уже касались выше общего метода *замены переменных*, или *подстановки* (нового неизвестного вместо первоначального). Если в интеграле $\int F(x)dx$ подынтегральная функция $F(x)$ может быть представлена в виде

$F(x)=f(\varphi(x))\varphi'(x),$

где $\varphi(x)$ — какая-то функция переменного x, то эту функцию можно обозначить одной буквой z и принять за новое неизвестное. Так как $\varphi'(x)dx=dz$, то мы придем при этом к интегралу $\int f(z)dz$, который может оказаться «табличным» (т. е. совпасть с одним из основных интегралов § 2; см. таблицу интегралов на с. 139—140) или быть проще первоначального. Этот общий метод используется в интегральном исчислении весьма широко.

Так, например, интегрирование многих функций, содержащих корни и тригонометрические функции, может

быть при помощи надлежащей замены переменной сведено к интегрированию многочленов или алгебраических дробей с целыми степенями. Рассмотрим несколько примеров.

Найти $\int x\sqrt{x+1}\,dx$. Сделаем замену переменной: $z=\sqrt{x+1}$, $x+1=z^2$. Отсюда $x=z^2-1$ и $dx=2zdz$. Переходя теперь в нашем интеграле от x к z, получаем

$$\int x\sqrt{x+1}\,dx=\int (z^2-1)z\cdot 2zdz=$$
$$=2\int (z^4-z^2)dz=2\frac{z^5}{5}-2\frac{z^3}{3}+C=$$
$$=\frac{2}{5}\sqrt{(x+1)^5}-\frac{2}{3}\sqrt{(x+1)^3}+C.$$

В интеграле $\int\frac{dx}{a^2+x^2}$ уместна замена $x=at$ $\left(\text{т. е. } t=\frac{x}{a}\right)$; при этом $dx=adt$ и мы имеем

$$\int\frac{dx}{a^2+x^2}=\int\frac{adt}{a^2+a^2t^2}=\int\frac{adt}{a^2(1+t^2)}=$$
$$=\frac{1}{a}\int\frac{dt}{1+t^2}=\frac{1}{a}\operatorname{arc\,tg} t+C=$$
$$=\frac{1}{a}\operatorname{arc\,tg}\frac{x}{a}+C.$$

В интеграле $\int\frac{dx}{\sqrt{(a^2-x^2)^3}}$ уместно положить $x=a\cos t$ $\left(\text{т. е. } t=\arccos\frac{x}{a}\right)$; при этом $dx=-a\sin t\,dt$, $a^2-x^2=a^2(1-\cos^2 t)=a^2\sin^2 t$ и мы получаем

$$\int\frac{dx}{\sqrt{(a^2-x^2)^3}}=\int\frac{-a\sin tdt}{\sqrt{(a^2\sin^2 t)^3}}=$$
$$=-\int\frac{a\sin tdt}{a^3\sin^3 t}=-\frac{1}{a^2}\int\frac{dt}{\sin^2 t}=$$
$$=\frac{1}{a^2}\operatorname{ctg} t+C=\frac{1}{a^2}\operatorname{ctg}\left(\arccos\frac{x}{a}\right)+C.$$

Это выражение можно еще упростить (см. упр. 10).

Почти та же подстановка применима и к «почти табличному» интегралу $\int\frac{dx}{\sqrt{a^2-x^2}}$ (который подстановкой $x=az$, $dx=adz$ сразу сводится к известному уже нам результату $\int\frac{dz}{\sqrt{1-z^2}}=\arcsin z+C$). В самом деле, если $x=a\sin t$, $dx=a\cos t\,dt$ и $\sqrt{a^2-x^2}=a\cos t$, то

$$\int\frac{dx}{\sqrt{a^2-x^2}}=\int\frac{a\cos t\,dt}{a\cos t}=\int dt=t+C=$$
$$=\arcsin\frac{x}{a}+C$$

$\left(\text{ибо } t=\arcsin\frac{x}{a}\right)$.

Аналогично может быть вычислен и интеграл $\int\frac{dx}{\sqrt{a^2+x^2}}$, только вместо тригонометрических функций здесь уместно использовать так называемые *гиперболический синус* $\operatorname{sh} t=\frac{1}{2}(e^t-e^{-t})$ и *гиперболический косинус* $\operatorname{ch} t=\frac{1}{2}(e^t+e^{-t})$ (о них см. § 14. 4). Легко видеть, что (проверьте это!) $\operatorname{ch}^2 t-\operatorname{sh}^2 t=1$, $(\operatorname{sh} t)'=\operatorname{ch} t$, $(\operatorname{ch} t)'=\operatorname{sh} t$. Поэтому если положить $x=a\operatorname{sh} t$, то $dx=a\operatorname{ch} t\,dt$ и $\sqrt{a^2+x^2}=a\operatorname{ch} t$. Таким образом, наш интеграл принимает вид

$$\int\frac{dx}{\sqrt{a^2+x^2}}=\int\frac{a\operatorname{ch} tdt}{a\operatorname{ch} t}=\int dt=t+C=$$
$$=\operatorname{arsh}\left(\frac{x}{a}\right)+C,$$

где «обратная гиперболическая» функция $\operatorname{arsh} z$ определяется из условия: если $t=\operatorname{arsh} z$, то $z=\operatorname{sh} t$.

Доведение этого примера до конца мы предоставляем читателю (см. упр. 11).

Наконец, приведем пример интеграла, который не может быть представлен при помощи конечной комбинации элементарных функций:

$$f(x)=\int e^{-x^2}dx. \tag{5.5.1}$$

Доказательство того, что функцию (1) нельзя выразить через известные нам «элементарные» функции (степенные, в том числе с дробными показателями, показательную, логарифмическую, тригонометрические и обратные тригонометрические), весьма сложно, и мы его приводить не будем.

Интеграл (1) является новой функцией, свойства которой можно изучить. Из определения $f(x)$ следует, что

$$\frac{df}{dx}=e^{-x^2}. \tag{5.5.2}$$

Так как $f'(x)=e^{-x^2}>0$ при любом x, то $f(x)$ — возрастающая функция. Производная функции $f(x)$ максимальна при $x=0$, где она равна 1; значит, график $f(x)$ образует наибольший угол с осью x при $x=0$ (этот угол равен 45°).

При больших по абсолютной величине (положительных и отрицательных) значениях x производная $\frac{df}{dx}$ очень мала; поэтому здесь функция $f(x)$ почти постоянна. График функции $\Phi(x)=\int\limits_0^x e^{-x^2}dx$ изображен на рис. 1 (для определенности нижний предел выбран равным нулю). Тот факт, что число $\Phi(\infty)\approx 0{,}885$ равно $\frac{\sqrt{\pi}}{2}$, в настоящей книге мы приводим без доказательства.

Для функции $\Phi(x)$ составлены подробные таблицы; благодаря им вычисления, в которых фигурирует интеграл (1), не сложнее, чем, например, расчеты с применением тригонометрических функций.

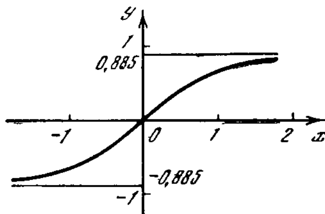


Рис. 5.5.1

Заметим еще, что, пользуясь различными приемами, иногда можно получить для одного и того же интеграла разные выражения (ср. упр. 6). Это не должно смущать читателя. Если вычисления сделаны верно, то полученные выражения должны отличаться лишь на постоянную — и при нахождении определенного интеграла (при любых пределах интегрирования) результаты окажутся одинаковыми.

Мы видим, что вычисление интегралов заметно сложнее нахождения производных: здесь приходится пользоваться разными искусственными приемами (интегрирование по частям, замена переменных) — и часто сразу не ясно, какой прием следует применить (например, как выбрать новую переменную). Более того, в некоторых случаях (как в случае интеграла (1)) вообще никакой прием не позволяет довести задачу до конца (т. е. выразить интеграл через конечное число элементарных функций); при этом мы не имеем никакого критерия, позволяющего судить о том, «берется» ли данный интеграл или «не берется»[1]. Технику интегрирования невозможно довести до того уровня автоматизма, с каким мы дифференцируем функции (т. е. находим их производные). При необходимости часто искать интегралы функций естественно пользоваться специальными справочниками, в которых имеются таблицы (неопределенных) интегралов, классифицированные по типам подынтегральных функций так, чтобы требуемый интеграл было легко найти[2], а также некоторые типы определенных интегралов (в пределах от 0 до ∞, от 0 до π, от $-\infty$ до ∞ и т. д.). В случаях, когда интеграл найти не удается, приходится прибегать к численному (приближенному) интегрированию.

Упражнения

Найдите интегралы:

5.5.1. $\int\cos(3x-5)\,dx$. **5.5.2.** $\int\sin(2x+1)\,dx$. **5.5.3.** $\int\sqrt{3x-2}\,dx$. **5.5.4.** $\int\frac{x\,dx}{x+\sqrt{x}}$. [Указание. Сделайте замену переменной $\sqrt{x}=z$.] **5.5.5.** $\int\frac{x\,dx}{\sqrt{x^2-5}}$. [Указание.

---

[1] Нередки случаи, когда формула для определенного интеграла $\int\limits_a^b f(x)\,dx$ в известных пределах от a до b существует, а общую формулу для неопределенного интеграла от той же функции найти не удается. Так, мы уже отмечали, что $\int\limits_0^\infty e^{-x^2}dx=\frac{\sqrt{\pi}}{2}$, где $\int\limits_0^\infty e^{-x^2}dx=\lim\limits_{M\to\infty}\int\limits_0^M e^{-x^2}dx$, в то время как соответствующий неопределенный интеграл вычислен быть не может. (Об одном из методов вычисления определенных интегралов, не требующих предварительного нахождения формулы для соответствующего неопределенного интеграла, рассказывается в § 17.3.)

[2] См., например, книги: *Бронштейн И. Н.*, *Семендяев К. А.* Справочник по математике для инженеров и учащихся вузов. М.: Наука, 1967; *Смолянский М. Л.* Таблицы неопределенных интегралов. М.: Наука, 1967; *Двайт Г.* Таблицы интегралов и других математических формул. М.: Изд-во иностр. лит., 1948; *Градштейн И. С.*, *Рыжик И. М.* Таблицы интегралов, сумм, рядов и произведений. М.; Л.: Гостехиздат, 1951. Последняя книга является наиболее полной.

Сделайте замену переменной $x^2 - 5 = z$.] 5.5.6. $\int \sin^3 x \cos x dx$. [Указание. Сделайте замену переменной $\sin x = z$ или $\cos x = u$.] 5.5.7. $\int \frac{\cos^3 x}{\sin^4 x} dx$. [Указание. Сделайте замену переменной $\sin x = z$.] 5.5.8. $\int \operatorname{tg} x dx$. 5.5.9. $\int \frac{dx}{\sqrt{a^2 - x^2}}$.

5.5.10. Докажите, что $\int \frac{dx}{\sqrt{(a^2 - x^2)^3}} = \frac{x}{a^2\sqrt{a^2 - x^2}} + C$.

5.5.11. Найдите интеграл $\int \frac{dx}{\sqrt{a^2 + x^2}}$.

5.5.12. Найдите интеграл $\int \frac{dx}{(x^2 + a^2)^2}$.

§ 6. Замена переменной в определенном интеграле

Рассмотрим пример. Пусть требуется вычислить определенный интеграл

$$\int_p^q (ax + b)^2 dx.$$

Можно поступить следующим образом: вычислить сначала неопределенный интеграл $\int (ax + b)^2 dx$, а затем составить разность его значений при $x = q$ и $x = p$.

Для вычисления $\int (ax + b)^2 dx$ сделаем замену переменной по формуле $z = ax + b$. Тогда $dz = adx$ и

$$\int (ax + b)^2 dx = \frac{1}{a} \int z^2 dz = \frac{z^3}{3a} = \frac{(ax + b)^3}{3a}.$$

Поэтому

$$\int_p^q (ax + b)^2 dx = \left.\frac{(ax + b)^3}{3a}\right|_p^q = \frac{(aq + b)^3 - (ap + b)^3}{3a}.$$

Можно, однако, действовать иначе. Выясним, как будет изменяться z, когда x изменяется от p до q. Так как z связано с x формулой $z = ax + b$, то при изменении x от p до q величина z будет меняться от $ap + b$ до $aq + b$. Следовательно,

$$\int_p^q (ax + b)^2 dx = \frac{1}{a} \int_{ap+b}^{aq+b} z^2 dz = \left.\frac{z^3}{3a}\right|_{ap+b}^{aq+b} = \frac{(aq + b)^3 - (ap + b)^3}{3a}.$$

При нахождении интегралов удобно поступать именно так: выполняя замену переменной, одновременно находить и новые пределы интегрирования. Тогда в выражении неопределенного интеграла не придется возвращаться к старой переменной.

Рассмотрим примеры.

1. Вычислим интеграл $\int_0^1 \frac{dx}{(2 - x)^3}$.

Заметим сразу же, что функция $\frac{1}{(2 - x)^3}$ при изменении x от 0 до 1 принимает положительные значения; поэтому $\int_0^1 \frac{dx}{(2 - x)^3} > 0$; вместе с тем знаменатель в этом промежутке не обращается в нуль, так что подынтегральная функция во всем промежутке конечна. Сделаем замену переменной $2 - x = y$, $dx = -dy$. Тогда $y = 2$ при $x = 0$; $y = 1$ при $x = 1$, так что

$$\int_0^1 \frac{dx}{(2 - x)^3} = -\int_2^1 \frac{dy}{y^3}. \tag{5.6.1}$$

В правой части (1) пределы интегрирования даны уже для y. Читателя может смутить знак минус в последнем равенстве. Действительно, справа и слева стоят интегралы от положительных функций; почему же положительна правая часть равенства (1)? Все дело в том, что в интеграле справа нижний предел больше верхнего. Так как при перестановке пределов интегрирования интеграл меняет знак, то равенство (1) можно записать так:

$$\int_0^1 \frac{dx}{(2 - x)^3} = \int_1^2 \frac{dy}{y^3}.$$

Теперь в интеграле справа верхний предел больше нижнего и ясно, что этот интеграл положителен. Вычисления легко доводятся до конца:

$$\int_1^2 \frac{dy}{y^3} = -\left.\frac{1}{2y^2}\right|_1^2 = -\frac{1}{8} + \frac{1}{2} = \frac{3}{8}.$$

2. В § 5 мы рассматривали функцию $\Phi(x) = \int_0^x e^{-x^2} dx$. Часто приходится иметь дело с функцией $\varphi(a) = \int_0^a e^{-kx^2} dx$, где

k — постоянное число. Покажем, что между функциями φ и Φ существует простая зависимость.

В выражении для $\varphi(a)$ сделаем замену переменной по формуле $kx^2=t^2$. Отсюда находим $x\sqrt{k}=t$, $x=\frac{t}{\sqrt{k}}$, $dx=\frac{1}{\sqrt{k}}dt$. Ясно, что $t=0$ при $x=0$; $t=a\sqrt{k}$ при $x=a$. Поэтому

$$\varphi(a)=\int_0^{a\sqrt{k}} e^{-t^2}\frac{dt}{\sqrt{k}}=\frac{1}{\sqrt{k}}\int_0^{a\sqrt{k}} e^{-t^2}dt=$$
$$=\frac{1}{\sqrt{k}}\Phi\left(a\sqrt{k}\right).$$

Следовательно, для любого значения независимой переменной x

$$\varphi(x)=\frac{1}{\sqrt{k}}\Phi\left(x\sqrt{k}\right). \qquad (5.6.2)$$

Имея таблицу значений функции $\Phi(x)$, можно найти также интеграл $\varphi(x)$ при любом значении k. И наоборот, зная значения функции $\varphi(x)$, отвечающие какому-либо конкретному значению k, мы можем с помощью зависимости (2) найти и значение функции $\Phi(z)$ для любого z (в (2) нам придется положить $z=x\sqrt{k}$)[1].

3. В гл. 3 мы видели, что если подынтегральная функция и пределы интегрирования имеют размерность, то и определенный интеграл не безразмерен. Часто, однако, бывает удобно привести интеграл к *безразмерному* виду, вынося за знак интеграла все размерные множители; при этом далее приходится рассматривать (преобразовывать) только численный (безразмерный) интеграл. Покажем, как это можно сделать.

Пусть дан интеграл $\int_a^b f(x)\,dx$, где функцию $f(x)$ мы для простоты считаем неотрицательной. Обозначим через f_{max} наибольшее значение функции $f(x)$ на промежутке интегрирования и вынесем число f_{max} за знак интеграла[2]:

$$\int_a^b f(x)\,dx=\int_a^b \frac{f(x)}{f_{max}}f_{max}dx=$$
$$=f_{max}\int_a^b \frac{f(x)}{f_{max}}dx. \qquad (5.6.3)$$

Ясно, что в последнем интеграле подынтегральная функция $\frac{f(x)}{f_{max}}$ уже безразмерна, так как $f(x)$ и f_{max} имеют одну и ту же размерность.

Перейдем теперь к безразмерному аргументу функции и безразмерным пределам интегрирования. Для этого выполним замену переменной по формуле

$$z=\frac{x-a}{b-a}, \text{ или } x=a+z(b-a). \qquad (5.6.4)$$

Из (4) видно, что величина z безразмерна. Так как $dx=(b-a)\,dz$ и $z=0$ при $x=a$, $z=1$ при $x=b$, то интеграл в правой части равенства (3) принимает вид

$$\int_a^b \frac{f(x)}{f_{max}}dx=(b-a)\int_0^1 \frac{f[a+z(b-a)]}{f_{max}}dz. \qquad (5.6.5)$$

Положим $\frac{f[a+z(b-a)]}{f_{max}}=\varphi(z)$; тогда из (5) следует

$$\int_a^b \frac{f(x)}{f_{max}}dx=(b-a)\int_0^1 \varphi(z)\,dz,$$

и окончательно получаем

$$\int_a^b f(x)\,dx=f_{max}(b-a)\int_0^1 \varphi(z)\,dz. \qquad (5.6.6)$$

В формуле (6) $\int_0^1 \varphi(z)\,dz$ — безразмерная величина.

---

[1] Заметим, что более привычными являются не таблицы значений функции $\Phi(x)=\int_0^x e^{-x^2}dx$, о которых мы говорили на с. 146, а таблицы значений родственного $\Phi(x)$ интеграла $\varphi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-x^2/2}dx$; однако функции $\varphi(x)$ и $\Phi(x)$ тесно связаны между собой (ср. формулу (2), где в этом случае следует положить $k=1/2$). Функция $\varphi(x)$ играет большую роль в одном из более продвинутых разделов высшей математики — в *теории вероятностей*, о которой упоминается в Заключении к этой книге.

[2] Мы положили, что $f(x)\geqslant 0$ и $f_{max}\neq 0$; в случае знакопеременной (или отрицательной) функции роль f_{max} должно играть наибольшее значение $|f|_{max}$ абсолютной величины функции f.

Если $f(x)$ мало меняется на промежутке интегрирования, то $\frac{f(x)}{f_{max}} \approx 1$ $\left(\text{но, разумеется, } \frac{f(x)}{f_{max}} \leqslant 1, \text{ ибо } f(x) \leqslant \leqslant f_{max}\right)$; в этом случае $\varphi(z) \approx 1$ (но $\varphi(z) \leqslant 1$) и $I = \int\limits_0^1 \varphi(z)\,dz \approx 1 \int\limits_0^1 dz = 1$ (но $I \leqslant 1$). А тогда безразмерный множитель $\int\limits_0^1 \varphi(z)\,dz = I$ имеет порядок величины 1 и интеграл (3) определяется в первую очередь произведением $f_{max}(b-a)$.

Рассмотрим элементарный пример: свободное падение тела в течение времени t_0. Путь в данном случае равен $\int\limits_0^{t_0} v(t)\,dt$. Ускорение тела равно g, и поэтому скорость тела $v(t) = gt$ (мы считаем, что $v(0) = 0$ и что максимальная скорость v_{max} достигается к моменту t_0: $v_{max} = gt_0$). Заметим, что здесь максимум обусловлен не тем, что после $t = t_0$ скорость начинает уменьшаться, а просто тем, что значения времени, которым должна была бы отвечать бóльшая, чем gt_0, скорость, не входят в промежуток $0 \leqslant t \leqslant t_0$ интегрирования. Вводим обозначения: $z = \frac{t}{t_0}$, $\varphi(z) = \frac{v}{v_{max}} = \frac{gt}{gt_0} = \frac{t}{t_0} = z$; тогда получаем

$$s = \int\limits_0^{t_0} v(t)\,dt = v_{max} t_0 \int\limits_0^1 z\,dz = gt_0^2 \int\limits_0^1 z\,dz = \frac{1}{2} gt_0^2.$$

Значит, в данном случае безразмерный множитель $I = \int\limits_0^1 \varphi(z)\,dz$ (значение которого в общем случае будет порядка 1, но который всегда меньше 1) равен 0,5, что достаточно хорошо соответствует нашей грубой прикидке.

Глава 6

РЯДЫ. ПРОСТЕЙШИЕ ДИФФЕРЕНЦИАЛЬНЫЕ УРАВНЕНИЯ

§ 1. Представление функций в виде рядов

Пусть функция $y(x)$ задана какой-либо формулой, которая, однако, кажется нам слишком сложной и неудобной для вычисления значений y. Поставим себе задачу — построить простое и удобное приближенное выражение функции $y(x)$ на небольшом промежутке изменения аргумента x, например при значениях x, близких к фиксированному числу a.

Определение производной (см. гл. 2) можно записать так:

$$y'(a)=\lim_{x\to a}\frac{y(x)-y(a)}{x-a},$$

откуда следует, что

$$y(x)\approx y(a)+(x-a)\,y'(a), \qquad (6.1.1)$$

где приближенное равенство (1) является «точным в пределе», т. е. будет тем точнее, чем меньше разность $|x-a|$. Формула (1) указывает, что при x, близком к a, изменению аргумента на (малую) величину $x-a$ отвечает изменение значения функции на $y'(a)(x-a)$, т. е. она согласуется со смыслом производной $\left.\frac{dy}{dx}\right|_{x=a}=y'(a)$ как *скорости изменения функции* в точке $x=a$.

При больших значениях $|x-a|$ формула (1) становится неточной: чем больше $|x-a|$, тем хуже она работает. Действительно, полагая $\Delta y=y(x)-y(a)=y'(a)(x-a)$, мы принимаем скорость изменения функции всюду между x и a равной скорости $y'(a)$ изменения функции в точке a, в то время как на этом промежутке изменения x сама скорость y' также меняется. Точная формула, учитывающая это непостоянство скорости $y'(t)$ изменения функции, имеет вид [1]

$$y(x)=y(a)+\int_a^x y'(t)\,dt. \qquad (6.1.2)$$

Формула (1) называется *первым*, или *линейным* [2], *приближением* функции $y(x)$.

Применим теперь формулу (1) к производной $y'(x)$. Мы получим

$$y'(x)\approx y'(a)+(x-a)\,y''(a). \qquad (6.1.3)$$

Прежде чем идти дальше, напомним читателю, что $y''(x)=\frac{dy'}{dx}$ — это *вторая производная* функции y по x, обозначаемая также $\frac{d^2y}{dx^2}$; она получается из y' так же, как y' — из y. Аналогично определяется $y'''=\frac{dy''}{dx}$ — третья производная функции y, ее четвертая производная y^{IV}, пятая производная y^{V} (или $y^{(5)}$) и т. д. *Производная n-го порядка*, получаемая в результате n-кратного последовательного дифференцирования функции y (n-кратного взятия производной), обозначается через $y^{(n)}$ или $\frac{d^ny}{dx^n}$ (но вторая, третья и четвертая производные всегда записываются как y'', y''' и y^{IV}, а не как $y^{(2)}$, $y^{(3)}$ и $y^{(4)}$). Скобки вверху в обозначении $y^{(n)}$ ставятся для того, чтобы отличить порядок производной от показателя степени.

Ясно, что если x измеряется в единицах e_1, а y — в единицах e_2, то размерность n-й производной $\frac{d^ny}{dx^n}$ есть $\frac{e_2}{(e_1)^n}$. Так, если y — путь (измеряемый в см), а x — время (в с), то вторая производная (ускорение!) y'' имеет размерность см/с^2; приращение

---

[1] Формула (2) непосредственно вытекает из (3.4.5а): ведь $\int_a^x y'(t)\,dt=y(t)\big|_a^x=y(x)-y(a)$. (Так же проверяется и формула (5).)

[2] В выражение (1) для y величина x входит только в первой степени, т. е. здесь y — многочлен первой степени от x. Такая зависимость графически изображается прямой линией (см. § 1.3) и потому называется *линейной*.

$\Delta y''$ ускорения имеет ту же размерность см/с^2, а, значит, размерность «скорости изменения ускорения» $y''' = \lim \frac{\Delta y''}{\Delta x}$ будет уже см/с^3 и т. д.

Вернемся к задаче приближенного выражения функции. Формула (3) для производной есть не что иное, как формула (1), примененная к функции y' (вместо y). Подставим (приближенное) выражение (3) для производной y' в точную формулу (2). Мы получим

$$y(x) \approx y(a) + \int_a^x [y'(a) + (t-a) y''(a)]\, dt = y(a) + (x-a) y'(a) + \frac{(x-a)^2}{2} y''(a). \tag{6.1.4}$$

Формула (4) тоже является приближенной, но она точнее выражения (1). При выводе (4) мы считались с тем, что производная $y'(x)$ непостоянна, но изменение $y'(x)$ учитывалось лишь приближенно: формула (3), которой мы пользовались в ходе вывода (4), получена в предположении постоянства $y''(x) (= y''(a))$, что приводит к л и н е й н о й зависимости y' от x; при этом для $y(x)$ получается к в а д р а т и ч н а я зависимость от x. Проведем еще для формулы (4) «контроль размерностей», ограничиваясь для простоты эталонным примером время — путь: если y измеряется в см, а x — в с, то размерности производных y' и y'' будут соответственно см/с (скорость) и см/с^2 (ускорение); величины $y(x)$ и $y(a)$ будут иметь размерность см, а множители $(x-a)$ и $\frac{1}{2}(x-a)^2$ — соответственно размерности с и с^2; поэтому размерности левой и правой частей формулы (4) будут одинаковы (см).

Уточним теперь формулу (4), учитывая, что и величина y'' непостоянна. Воспользуемся формулой

$$y'(x) = y'(a) + \int_a^x y''(t)\, dt, \tag{6.1.5}$$

которая получается из равенства (2) заменой y на y'. Представим $y''(x)$ формулой типа (1), только примененной уже не к исходной функции $y(x)$, а к функции $y''(x)$:

$$y''(x) \approx y''(a) + (x-a) y'''(a), \tag{6.1.6}$$

и подставим полученное выражение в (5). Из формул (5) и (6) следует

$$y'(x) \approx y'(a) + \int_a^x [y''(a) + (t-a) y'''(a)]\, dt,$$

или

$$y'(x) \approx y'(a) + (x-a) y''(a) + \frac{(x-a)^2}{2} y'''(a). \tag{6.1.7}$$

(Ясно, что (7) есть формула типа (4), записанная для функции $y'(x)$.)

Выражение для $y'(x)$ из (7) подставим в (2):

$$y(x) \approx y(a) + \int_a^x \left[y'(a) + y''(a)(t-a) + y'''(a) \frac{(t-a)^2}{2} \right] dt = y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \frac{y'''(a)}{6}(x-a)^3. \tag{6.1.8}$$

Общий закон усматривается из сравнения полученных нами выражений. В самом грубом приближении можно считать, что при малом $|x-a|$

$$y(x) \approx y(a) \tag{6.1.9}$$

(очевидное приближение (9) вообще не требует знания математики). Назовем (9) *нулевым приближением* функции $y(x)$, выражение (1) — *первым приближением*, (4) — *вторым приближением*, (8) — *третьим приближением* и выпишем все эти формулы рядом:

$y(x) \approx y(a)$ (нулевое приближение);

$y(x) \approx y(a) + (x-a) y'(a)$ (первое приближение);

$y(x) \approx y(a) + (x-a) y'(a) + \frac{(x-a)^2}{2} y''(a)$ (второе приближение);

$y(x) \approx y(a) + (x-a) y'(a) + \frac{(x-a)^2}{2} y''(a) + \frac{(x-a)^3}{6} y'''(a)$ (третье приближение).

Теперь легко представить себе, какой вид будут иметь приближенные формулы для $y(x)$, получаемые, если продолжать процесс их уточнения: так, если учесть, что y''' непостоянна, то в формулу войдет $y^{IV}(a)$; при этом выражение для $y(x)$ (четвертое приближение) будет содержать член

с $(x-a)^4$. Каждый следующий шаг по уточнению $y(x)$ дает дополнительный член с более высокой степенью $x-a$. При этом можно ожидать, что чем больше степеней $x-a$ входит в формулу, тем она будет точнее. Конечно, сказанное будет верно лишь при малых значениях $|x-a|$, а при больших $|x-a|$ все эти формулы могут оказаться неверными (ср. § 3 ниже).

Общий вид n-го приближения

$$y(x) \approx y(a) + c_1 y'(a)(x-a) + c_2 y''(a)(x-a)^2 + c_3 y'''(a)(x-a)^3 + \ldots + c_n y^{(n)}(a)(x-a)^n \quad (6.1.10)$$

вытекает и из соображений размерности. В самом деле, если y измеряется в единицах e_2, а x — в единицах e_1, то $y^{(k)}(a) = \left.\frac{d^k y}{dx^k}\right|_{x=a}$ имеет размерность $\frac{e_2}{(e_1)^k}$, так что в приближенном выражении для $y(x)$ слагаемое с $y^{(k)}(a)$ обязательно должно содержать множитель размерности $(e_1)^k$, т. е. множитель $(x-a)^k$. Правда, эти подсчеты не позволяют судить о численных значениях коэффициентов $c_1, c_2, c_3, \ldots$ в формуле (10) (мы знаем, что $c_1=1$, $c_2=1/2$, $c_3=1/6$, — но каковы c_4 и последующие коэффициенты?), однако эти величины можно определить следующим образом.

Обозначим правую часть формулы (10) (многочлен n-й степени от переменной x) через $P_n(x)$:

$$P_n(x) = y(a) + c_1 y'(a)(x-a) + c_2 y''(a)(x-a)^2 + c_3 y'''(a)(x-a)^3 + \ldots + c_n y^{(n)}(a)(x-a)^n. \quad (6.1.11)$$

Из (11), разумеется, следует, что $P_n(a) = y(a)$. Далее, дифференцирование равенства (11) дает

$$P'_n(x) = c_1 y'(a) + 2c_2 y''(a)(x-a) + 3c_3 y'''(a)(x-a)^2 + \ldots + nc_n y^{(n)}(a)(x-a)^{n-1}. \quad (6.1.11\text{a})$$

Поэтому $P'_n(a) = c_1 y'(a)$. Потребуем теперь, чтобы не только *значение* $P_n(a)$ *многочлена* (11) *в точке* $x=a$ *совпадало со значением* $y(a)$ *функции* $y(x)$ *в той же точке, но и чтобы скорость* $P'_n(a)$ *роста многочлена в точке* $x=a$ *не отличалась от скорости* $y'(a)$ *роста функции* $y(x)$. Для этого надо положить $c_1=1$. Аналогично, дифференцируя далее обе части равенства (11а) и подставляя в полученный результат значение $x=a$, мы последовательно получаем:

$$P''_n(a) = 2c_2 y''(a) \quad \text{и} \quad P''_n(a) = y''(a) \quad \text{при } c_2 = 1/2,$$

$$P'''_n(a) = 6c_3 y'''(a) \quad \text{и} \quad P'''_n(a) = y'''(a) \quad \text{при } c_3 = 1/6,$$

$$P^{IV}_n(a) = 24c_4 y^{IV}(a) \quad \text{и} \quad P^{IV}_n(a) = y^{IV}(a) \quad \text{при } c_4 = 1/24,$$

. . . .

$$P^{(n)}_n(a) = 2\cdot 3\cdot 4\cdot \ldots \cdot n y^{(n)}(a) \quad \text{и}$$

$$P^{(n)}_n(a) = y^{(n)}(a) \quad \text{при } c_n = 1/(2\cdot 3\cdot \ldots \cdot n).$$

Итак, если потребовать, чтобы *значения* n *последовательных производных многочлена* $P_n(x)$ *при* $x=a$ *совпадали со значениями производных того же порядка функции* $y(x)$ *в той же точке* (а это требование как раз и означает большую близость функций $P_n(x)$ и $y(x)$!), то мы должны будем положить $c_1=1$, $c_2=1/2$, $c_3=1/(2\cdot 3)$, $c_4=1/(2\cdot 3\cdot 4)$, ..., $c_n=1/(2\cdot 3\cdot \ldots \cdot n)$.

Таким образом,

$$P_n(x) = y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \ldots + \frac{y^{(n)}(a)}{2\cdot 3\cdot 4\cdot \ldots \cdot n}(x-a)^n. \quad (6.1.12)$$

Мы пришли к приближенной формуле

$$y(x) \approx y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 + \frac{y'''(a)}{6}(x-a)^3 + \ldots + \frac{y^{(n)}(a)}{2\cdot 3\cdot \ldots \cdot n}(x-a)^n. \quad (6.1.13)$$

Произведение $1\cdot 2\cdot 3\cdot \ldots \cdot n$ первых n последовательных натуральных чисел принято обозначать символом $n!$ (читается: «эн факториал», от латинского слова factor — сомножитель): так $3! = 1\cdot 2\cdot 3 = 6$, $4! = 1\cdot 2\cdot 3\cdot 4 = 3!\cdot 4 = 24$, $5! = 1\cdot 2\cdot 3\cdot 4\cdot 5 = 120$, а $1! = 1$ и $2! = 2$. В этих обозначениях формула (13) запишется так:

$$y(x) \approx y(a) + \frac{y'(a)}{1!}(x-a) + \frac{y''(a)}{2!}(x-a)^2 + \ldots + \frac{y^{(n)}(a)}{n!}(x-a)^n. \quad (6.1.13\text{a})$$

Вот более полное доказательство той же формулы (13) (или (13а)), приводящее к более сильным, чем ранее, результатам. Вернемся к точному равенству (2) и заменим под интегралом dt на $d(t-x)$:

$$y(x) = y(a) + \int_a^x y'(t)\,dt =$$
$$= y(a) + \int_a^x y'(t)\,d(t-x).$$

Эта замена законна, ибо здесь x (верхний предел интегрирования) рассматривается как постоянная, и потому $d(t-x)=dt$.

Воспользуемся теперь приемом интегрирования по частям (см. формулу (5.4.3а)):

$$y(x) = y(a) + \int_a^x y'(t)\,d(t-x) =$$
$$= y(a) + y'(t)(t-x)\Big|_a^x - \int_a^x (t-x)\,dy'(t) =$$
$$= y(a) + y'(a)(x-a) +$$
$$+ \int_a^x (x-t)\,y''(t)\,dt. \qquad (6.1.14)$$

Если теперь снова заменить под интегралом dt на $d(t-x)$ и проинтегрировать по частям, то получим

$$\int_a^x (x-t)\,y''(t)\,d(t-x) =$$
$$= -\int_a^x y''(t)\,d\frac{(x-t)^2}{2} = -y''(t)\,\frac{(x-t)^2}{2}\Big|_a^x +$$
$$+ \int_a^x \frac{(x-t)^2}{2}\,dy''(t) = y''(a)\,\frac{(x-a)^2}{2} +$$
$$+ \frac{1}{2}\int_a^x (x-t)^2\,y'''(t)\,dt,$$

и, значит,

$$y(x) = y(a) + y'(a)(x-a) + \frac{y''(a)}{2}(x-a)^2 +$$
$$+ \frac{1}{2}\int_a^x (x-t)^2\,y'''(t)\,dt. \qquad (6.1.14a)$$

Производя интегрирование по частям n раз, мы получим выражение, состоящее из $n+2$ членов:

$$y(x) = y(a) + y'(a)(x-a) +$$
$$+ \frac{y''(a)}{2!}(x-a)^2 + \frac{y'''(a)}{3!}(x-a)^3 + \ldots$$
$$\ldots + \frac{y^{(n)}(a)}{n!}(x-a)^n +$$
$$+ \frac{1}{n!}\int_a^x (x-t)^n\,y^{(n+1)}(t)\,dt. \qquad (6.1.15)$$

Формула (15) в противоположность (13а) является точной: ведь она получена преобразованием точной формулы (2)! Последний член $r_n = \frac{1}{n!}\int_a^x (x-t)^n\,y^{(n+1)}(t)\,dt$ в ее правой части называется *остатком*, или *остаточным членом*: он указывает разность между левой частью $y(x)$ формулы (13) и приближенно выражающей ее правой частью той же формулы.

Формулу (15) удобнее переписать в форме, не содержащей интеграла[3]:

$$y(x) = y(a) + y'(a)(x-a) +$$
$$+ \frac{y''(a)}{2!}(x-a)^2 + \ldots + \frac{y^{(n)}(a)}{n!}(x-a)^n +$$
$$+ \frac{(x-a)^{n+1}}{(n+1)!}\,y^{(n+1)}(c), \qquad (6.1.16)$$

где c — какое-то число, заключенное между x и a.

В общем случае произвольной функции $y(x)$ никакое конечное число степеней $(x-a)$, взятых с постоянными коэффициентами, не выразит точно[4] исходную функцию: правая часть (13) (или (13а)) лишь п р и б л и ж е н н о совпадает с $y(x)$, но не равна $y(x)$. Точную формулу можно получить, лишь суммируя б е с к о н е ч н о е множество степеней $(x-a)$:

$$y(x) = c_0 + c_1(x-a) + c_2(x-a)^2 +$$
$$+ c_3(x-a)^3 + \ldots + c_n(x-a)^n + \ldots \qquad (6.1.17)$$

---

[3] Относительно справедливости перехода от (15) к (16) см. текст, напечатанный в § 7.7 мелким шрифтом, в частности формулу (7.7.13), где надо только обозначить переменную интегрирования через $x-t$ $(=z)$ (теперь x постоянно, так что $d(x-t)=-dt$), а зависящую от новой переменной $x-t=z$ функцию $y^{(n+1)}(t)$ принять за $f(z)$. Тогда мы получим

$$\int_a^x (x-t)^n\,y^{(n+1)}(t)\,dt =$$
$$= \frac{(x-a)^{n+1}}{n+1}\,y^{(n+1)}(c),$$

где c — какое-то значение независимой переменной, промежуточное между x и a.

[4] Исключением здесь является случай, когда $y(x)$ — это многочлен (ср. с началом § 2).

Стоящее в правой части (17) выражение называют **бесконечным рядом** (впрочем, прилагательное «бесконечный» обычно опускают и говорят просто **ряд**). При этом про равенство (17) тоже можно сказать, что оно «точно в пределе», однако здесь этот оборот имеет несколько иной смысл, чем выше. Дело в том, что, разумеется, нельзя сложить сразу бесконечно много слагаемых; поэтому, выписывая стоящую справа сумму, мы неизбежно должны будем ограничиться тем или иным конечным числом n первых ее членов. Утверждение, что при достаточно малых $|x-a|$ формула (17) «точна», означает, что сумму справа можно сделать с к о л ь у г о д н о б л и з к о й к $y(x)$ (для этого только нужно взять достаточно большое число n слагаемых). Можно также сказать, что $y(x)$ равно п р е д е л у суммы n первых членов ряда при $n \to \infty$ [5].

Коэффициенты c_0, c_1, ..., c_n, ... в (17) различны для разных функций; они также зависят от значения a. Мы уже видели, что $c_0=y(a)$, $c_1=y'(a)$, $c_2=\frac{y''(a)}{2!}$, $c_3=\frac{y'''(a)}{3!}$ и т. д., т. е.

$$y(x)=y(a)+y'(a)(x-a)+\frac{y''(a)}{2!}(x-a)^2+\frac{y'''(a)}{3!}(x-a)^3+\ldots\ldots+\frac{y^{(n)}(a)}{n!}(x-a)^n+\ldots \quad (6.1.18)$$

Отметим еще частный случай формулы (18), получаемый в предположении, что $a=0$:

$$y(x)=y(0)+y'(0)x+\frac{y''(0)}{2}x^2+\frac{y'''(0)}{6}x^3+\ldots+\frac{y^{(n)}(0)}{n!}x^n+\ldots \quad (6.1.19)$$

С помощью символа суммирования формулы (18) и (19) можно сжато записать так:

$$y(x)=y(a)+\sum_{n=1}^{n=\infty}\frac{y^{(n)}(a)}{n!}(x-a)^n, \quad (6.1.18a)$$

$$y(x)=y(0)+\sum_{n=1}^{n=\infty}\frac{y^{(n)}(0)}{n!}x^n. \quad (6.1.19a)$$

Формулы (18) и (19) (или (18a) и (19a)) дают разложение функции $y(x)$ в ряд по степеням $x-a$ (или x). Стоящий в правой части формулы (18) (или (18a)) ряд называется *рядом Тейлора*; частный случай этого ряда, стоящий в правой части формулы (19) (или (19a)), называется *рядом Маклорена*.

Примерам приложений формул Тейлора и Маклорена к конкретным функциям будет посвящен следующий параграф, однако первое подобное представление функции в виде ряда мы рассмотрим уже здесь.

Пусть $y=e^x$. Тогда

$$y'=e^x, \quad y''=e^x, \ldots, \quad y^{(n)}=e^x, \ldots$$

Воспользуемся формулой (19) разложения функции в ряд Маклорена. В нашем случае

$$y(0)=y'(0)=y''(0)=\ldots=1.$$

Подставляя эти значения в (19), получим разложение функции $y=e^x$ в ряд Маклорена

$$e^x=1+x+\frac{x^2}{2}+\frac{x^3}{6}+\ldots\ldots+\frac{x^n}{n!}+\ldots \quad (6.1.20)$$

Рассмотрим теперь формулу, которая получается из формулы Тейлора (18), если ограничиться, например, тремя членами ряда:

$$y(x)\approx y(a)+(x-a)y'(a)+\frac{(x-a)^2}{2}y''(a).$$

Раскроем в правой части скобки и расположим результат по степеням x:

$$y(x)\approx\left[y(a)-ay'(a)+\frac{a^2}{2}y''(a)\right]+[y'(a)-ay''(a)]x+\frac{1}{2}y''(a)x^2. \quad (6.1.21)$$

Справа в (21) стоит многочлен второй степени. Обращаем внимание читателя на то, что этот многочлен не совпадает с тем, который получится, если взять три члена ряда Маклорена:

$$y(x)\approx y(0)+y'(0)x+\frac{y''(0)}{2}x^2. \quad (6.1.21a)$$

[5] При этом возможно, что при разных значениях x для достижения одной и той же точности приближения суммы первых членов ряда к значению $y(x)$ потребуется р а з н о е число слагаемых. (Мы не останавливаемся пока на общих вопросах об области применимости формул вида (17) и о функциях, для которых разложения (18) и (19) не верны.)

Это и понятно: ведь формула (21) дает хороший результат, если x близко к a, а формула (21а) точна при x, близком к нулю.

В гл. 2 мы определили производную как предел отношения приращения функции к приращению независимой переменной. После того как мы построили выражение функции в виде ряда, можно уточнить закон приближения отношения $\frac{\Delta y}{\Delta x}$ к производной $\frac{dy}{dx}$ при уменьшении Δx.

Возьмем ряд Тейлора (18) и обозначим $x-a$ через Δx. Затем перенесем $y(a)$ в левую часть (18), разделим обе части на $x-a=\Delta x$ и обозначим $y(x)-y(a)$ через Δy. Мы получим

$$\frac{\Delta y}{\Delta x}=y'(a)+\frac{1}{2}y''(a)\Delta x+\frac{1}{6}y'''(a)(\Delta x)^2+\ldots \qquad (6.1.22)$$

При малых Δx второй член справа больше третьего члена. Пренебрегая последним, придем к выводу, что отличие отношения $\frac{\Delta y}{\Delta x}$ от значения $y'(a)$ производной на краю промежутка от $x=a$ до $x=a+\Delta x$ пропорционально величине второй производной $y''(a)$ и длине промежутка Δx:

$$\frac{\Delta y}{\Delta x}\approx y'(a)+\frac{1}{2}y''(a)\Delta x.$$

При этом мы сравниваем отношение приращения функции на рассматриваемом участке значений x к приращению x с производной $y'(a)$ на к р а ю этого участка.

Производную можно вычислять и иначе, принимая за ее приближенное значение не дробь $\frac{f(x+\Delta x)-f(x)}{\Delta x}$, а дробь $\frac{f(x+\Delta x/2)-f(x-\Delta x/2)}{\Delta x}$. Другими словами, при вычислении производной в точке a мы составляем приращение, которое получит функция при изменении аргумента от $a-\frac{\Delta x}{2}$ до $a+\frac{\Delta x}{2}$, и делим на Δx именно это приращение. Сравним теперь отношение нашего приращения функции $y=f(x)$ к Δx с производной $y'(a)$, вычисленной в с е р е д и н е участка изменения переменной. Мы найдем, что

$$\Delta y=f\left(a+\frac{\Delta x}{2}\right)-f\left(a-\frac{\Delta x}{2}\right);$$

далее,

$$f\left(a+\frac{\Delta x}{2}\right)\approx f(a)+\frac{\Delta x}{2}f'(a)+\frac{1}{2}\left(\frac{\Delta x}{2}\right)^2 f''(a)+\frac{1}{6}\left(\frac{\Delta x}{2}\right)^3 f'''(a),$$

$$f\left(a-\frac{\Delta x}{2}\right)\approx f(a)-\frac{\Delta x}{2}f'(a)+\frac{1}{2}\left(\frac{\Delta x}{2}\right)^2 f''(a)-\frac{1}{6}\left(\frac{\Delta x}{2}\right)^3 f'''(a),$$

$$\frac{\Delta y}{\Delta x}\approx f'(a)+\frac{2}{6\cdot 2}\left(\frac{\Delta x}{2}\right)^2 f'''(a)=f'(a)+\frac{(\Delta x)^2}{24}f'''(a). \qquad (6.1.23)$$

Мы видим, что новый способ оценки производной $f'(a)$ гораздо точнее старого: разность между отношением приращений и производной теперь пропорциональна $(\Delta x)^2$, а не (Δx), и к тому же выражение для этой разности содержит малый коэффициент 1/24. Итак, мы имеем приближенно

$$f'(a)\approx\frac{f(a+\Delta x)-f(a)}{\Delta x}, \qquad (6.1.24)$$

а точнее, из формулы (22)

$$f'(a)\approx\frac{f(a+\Delta x)-f(a)}{\Delta x}-\frac{1}{2}f''(a)\Delta x. \qquad (6.1.24а)$$

Или, иначе,

$$f'(a)\approx\frac{f(a+\Delta x/2)-f(a-\Delta x/2)}{\Delta x}, \qquad (6.1.25)$$

а точнее (в силу формулы (23)),

$$f'(a)\approx\frac{f(a+\Delta x/2)-f(a-\Delta x/2)}{\Delta x}-\frac{1}{24}f'''(a)(\Delta x)^2. \qquad (6.1.25а)$$

При этом равенства (24а) и (25а) (как и равенства (24), (25)), разумеется, являются приближенными: в них мы пренебрегаем величинами более высокого порядка малости (по отношению к Δx), чем те, которые входят в эти формулы.

Аналогично можно получить и приближенные формулы для вычисления второй производной. По определению,

$$f''=\frac{df'}{dx}\approx\frac{\Delta f'}{\Delta x}.$$

Заменяя теперь $\Delta f'$ на $f'(a+\Delta x)-f'(a)$ и оценивая $f'(a+\Delta x)$ и $f'(a)$ с помощью отношений $\frac{f(a+2\Delta x)-f(a+\Delta x)}{\Delta x}$ и $\frac{f(a+\Delta x)-f(a)}{\Delta x}$, мы приходим к прибли-

женному выражению для второй производной функции, в которой фигурирует отношение $\frac{f(a+2\Delta x)-2f(a+\Delta x)+f(a)}{(\Delta x)^2}$. Для оценки этого отношения обратимся снова к формуле Тейлора:

$$f(a+2\Delta x)\approx f(a)+\frac{1}{2}f'(a)(2\Delta x)+\frac{1}{6}f''(a)(2\Delta x)^2+\frac{1}{24}f'''(a)(2\Delta x)^3,$$

$$f(a+\Delta x)\approx f(a)+\frac{1}{2}f'(a)\Delta x+\frac{1}{6}f''(a)(\Delta x)^2+\frac{1}{24}f'''(a)(\Delta x)^3.$$

Отсюда, учитывая первые три члена в правых частях этих формул, получим

$$\frac{f(a+2\Delta x)-2f(a+\Delta x)+f(a)}{(\Delta x)^2}\approx\frac{1}{3}f''(a),$$

или, точнее (с учетом всех четырех членов),

$$f''(a)\approx 3\frac{f(a+2\Delta x)-2f(a+\Delta x)+f(a)}{(\Delta x)^2}-\frac{3}{4}f'''(a)\Delta x. \tag{6.1.26}$$

Для получения $f''(a)$ лучше комбинировать не значения $f(a)$, $f(a+\Delta x)$, $f(a+2\Delta x)$

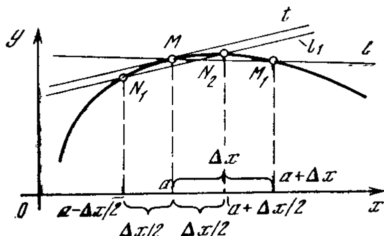


Рис. 6.1.1

функции, вычисленные в промежутке от a до $a+2\Delta x$, а значения $f(a-\Delta x)$, $f(a)$, $f(a+\Delta x)$. (Здесь интересующее нас значение $x=a$ совпадает с *серединой* области изменения аргумента.) И в самом деле, легко показать, что

$$f''(a)\approx 3\frac{f(a+\Delta x)-2f(a)+f(a-\Delta x)}{(\Delta x)^2},$$

или, точнее,

$$f''(a)\approx 3\frac{f(a+\Delta x)-2f(a)+f(a-\Delta x)}{(\Delta x)^2}-\frac{1}{20}f^{IV}(a)(\Delta x)^4 \tag{6.1.27}$$

(см. упр. 6).

Геометрически различия между приближениями (24) и (25) для производной сводятся к тому, что в первом случае мы заменяем касательную t к графику функции $y=f(x)$ (рис. 1) в точке $M(a, f(a))$ (наклон этой касательной равен производной $f'(a)$) прямой $MM_1\equiv l$, где $M_1=M_1(a+\Delta x,\ f(a+\Delta x))$, а во втором случае — прямой $N_1N_2=l_1$, где $N_1=N_1\left(a-\frac{\Delta x}{2},\ f\left(a-\frac{\Delta x}{2}\right)\right)$, а $N_2=N_2\left(a+\frac{\Delta x}{2}, f\left(a+\frac{\Delta x}{2}\right)\right)$. Но ясно, что прямая l_1 по направлению гораздо ближе к касательной t, чем прямая l (рис. 1). Аналогично формула (26) получается в предположении, что скорость изменения производной y' (скорость поворота касательной к кривой $y=f(x)$) оценивается по разностям изменений производной на участках от M до M_1 и от M_1 до M_2 $(x+2\Delta x, f(x+2\Delta x))$, а формула (27) — что скорость изменения y' оценивается по разностям изменений производной на участках от M' $(x-\Delta x, f(x-\Delta x))$ до M и от M до M_1 (чертеж сделайте сами).

Упражнения

6.1.1. Разложите многочлен 3-й степени $y=ax^3+bx^2+cx+d$ в ряд по степеням $(x-x_0)$, где x_0 — произвольное фиксированное число; сравните сумму первых двух, трех и четырех членов полученного разложения с многочленом.

6.1.2. Разложите в ряд Маклорена функцию $y=xe^x$; проверьте, что разложение этой функции можно получить из разложения функции $y=e^x$ умножением последнего на x.

6.1.3. Разложите функцию e^x в ряд Тейлора по степеням $(x-1)$.

6.1.4. Выясните точность приближенной формулы $(1+r)^m\approx e^{mr}$ при $r\ll 1$ и $mr\ll 1$, записав приближенно левую и правую части равенства в виде сумм $a+br+cr^2$, полученных в пренебрежении 3-й и высшими степенями r.

6.1.5. Используя формулы (24) и (25) (соответственно (26) и (27)) и задаваясь интервалом $\Delta x=1$; 1/2; 1/4; 1/8, найдите приближенные значения: а) производной функции e^x при $x=0$; б) второй производной функции e^x при $x=0$. Сравните полученные результаты с точными значениями.

6.1.6. Докажите формулу (27).

§ 2. Вычисление значений функций при помощи рядов

Остановимся вкратце на принципах, положенных в основу формул § 1. При первых подходах к понятию производной (т. е. скорости изменения функции) мы считали рассматриваемые функции известными, т. е. исходили из того, что можем вычислить значение

функции при любом значении аргумента. Эта установка позволяла нам находить производные непосредственно, так сказать опытным путем, вычисляя значения функции при значениях аргумента, близких к рассматриваемому нами. Потом мы научились находить производные по формулам, и оказалось, что составление формул для производных — довольно простое дело. Поэтому часто вычисление значений функций при помощи формул, в которые входят производные (а именно такозыми были все приближенные формулы § 1!), оказывается более простым, чем прямое вычисление значений функции.

Функции, отличные от многочлена, представляются бесконечными рядами, так как только в случае многочлена ряд Тейлора обрывается, содержит конечное число членов (это утверждение мы докажем чуть позже). Практическая ценность для вычислений бесконечных рядов связана с возможностью, ограничиваясь небольшим числом (зачастую двумя или тремя) членов ряда, получить достаточно точный результат. Для этого необходимо, разумеется, чтобы сумма отброшенных членов была достаточно мала.

Рассмотрим несколько примеров разложения функций в ряды. Мы уже упоминали (простейший) случай, когда функция

$$y = b_0 + b_1x + b_2x^2 + \ldots + b_nx^n \qquad (6.2.1)$$

есть многочлен степени n. Его производная $y'(x)$ будет многочленом степени $n-1$, $y''(x)$ — многочленом степени $n-2$ и т. д., $y^{(n)}(x)$ есть постоянная величина, а $y^{(n+1)}(x) \equiv 0$ и все производные более высокого порядка — также нули. Поэтому для многочлена ряд Тейлора (1.18) обрывается, он состоит из конечного числа членов. В качестве ряда Тейлора мы получаем тот же многочлен, только расположенный по степеням $(x-a)$:

$$y = c_0 + c_1(x-a) + c_2(x-a)^2 + \ldots$$
$$\ldots + c_n(x-a)^n. \qquad (6.2.1a)$$

При этом для многочленов n-й степени сумма первых $n+1$ членов ряда Тейлора дает точное равенство, верное при любых x, а не только при x, близких к a; само же выражение (1) многочлена можно понимать как формулу Маклорена (1. 19) для нашей функции $y(x)$ (почему?).

В предыдущем параграфе мы уже рассматривали экспоненциальную функцию и получили формулу (1. 20)

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots$$
$$\ldots + \frac{x^n}{n!} + \ldots \qquad (6.2.2)$$

В частности, подставляя сюда $x=1$ и $x=-1$, получим выражения для самого числа e и для обратного e числа $\frac{1}{e}$ в виде рядов:

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \ldots + \frac{1}{n!} + \ldots; \qquad (6.2.3)$$

$$\frac{1}{e} = 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} + \ldots$$
$$\ldots + \frac{1}{(2n)!} - \frac{1}{(2n+1)!} + \ldots \qquad (6.2.3a)$$

Формула (2) позволяет быстро и с большой точностью вычислять e^x, что можно видеть из табл. 1.

Если при $x=0{,}1$ мы ограничимся первыми двумя членами формулы, то

Таблица 1

x	e^x	$1+x$	$1+x+\frac{x^2}{2}$	$1+x+\frac{x^2}{2}+\frac{x^3}{6}$	$1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}$
0,10	1,1052	1,10	1,1050	1,1052	1,1052
0,25	1,2840	1,25	1,2812	1,2838	1,2840
0,50	1,6487	1,50	1,6250	1,6458	1,6484
0,75	2,1170	1,75	2,0312	2,1015	2,1147
1,00	2,7183	2,00	2,5000	2,6667	2,7083
1,25	3,4903	2,25	3,0312	3,3568	3,4585
1,50	4,4817	2,50	3,6250	4,1876	4,3986
2,00	7,3891	3,00	5,0000	6,3333	7,0000

получим ошибку, не превосходящую 0,5% истинного значения функции. При $x=0,5$ первые три члена формулы дают ошибку в 1,4%. При $x=1,0$ первые четыре члена формулы дают ошибку в 1,8% значения функции $y=e^x$.

Такая высокая точность связана с тем, что члены ряда (2) убывают. Каждый член этого ряда меньше предшествующего ему прежде всего потому, что знаменатель $(n+1)$-го члена в n раз больше знаменателя предыдущего (n-го) члена ряда. А если $x<1$,

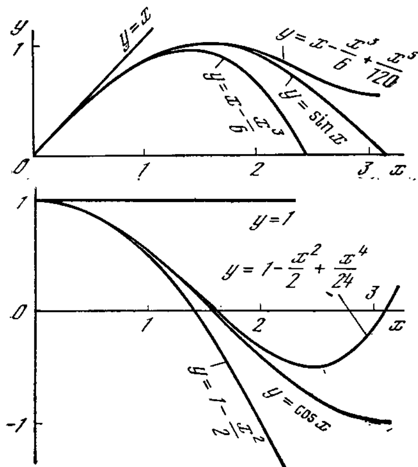


Рис. 6.2.1

то к этому добавляется еще и то обстоятельство, что с ростом n величина x^n быстро убывает. Но даже и при $x>1$ в далеких от начала участках ряда увеличение знаменателя с ростом n обязательно пересиливает рост числителя: ведь числитель при переходе от n-го члена к $(n+1)$-му увеличивается в x раз, а знаменатель возрастает в n раз — и нам надо только дождаться, пока (при движении вдоль ряда) n не станет больше x. Как видно из табл. 1, при $x=2$ сумма пяти членов ряда дает ошибку 5%; если же добавить шестой член $\left(\frac{x^5}{120}\right)$, то сумма станет равной 7,3500 — здесь ошибка составит всего 0,5%.

Построим формулы такого же типа для тригонометрических функций. Так, для $y(x)=\sin x$ имеем:

$$y'(x)=\cos x, \quad y''(x)=-\sin x,$$

$$y'''(x)=-\cos x, \quad y^{IV}(x)=\sin x.$$

Закон образования последующих производных очевиден.

Подставляя $x=0$, получим:

$$y(0)=0, \quad y'(0)=1, \quad y''(0)=0,$$

$$y'''(0)=-1, \ \ldots$$

Следовательно,

$$\sin x = x-\frac{x^3}{6}+\frac{x^5}{120}-\frac{x^7}{5040}+\ldots \tag{6.2.4}$$

Аналогично получим формулу

$$\cos x = 1-\frac{x^2}{2}+\frac{x^4}{24}-\frac{x^6}{720}+\ldots \tag{6.2.5}$$

На рис. 1 показаны графики синуса, косинуса, а также графики многочленов, которые получаются, если брать один, два или три члена соответствующего ряда. Видно, как улучшается точность по мере того, как мы берем все большее число членов ряда.

В табл. 2 и 3 приведены численные значения синуса и косинуса соответственно; в этих таблицах φ, как и x, — угол, но выраженный не в радианах, как x, а в градусах. Как видно из таблиц, достаточно двух-трех членов ряда, чтобы получить в промежутке от 0 до $\frac{\pi}{4}$ прекрасную точность. Таким образом, степенной ряд дает очень удобный практический способ вычисления значений тригонометрических функций. Заметим, что по абсолютной величине не равные нулю члены ряда для синуса и косинуса в точности равны соответствующим членам ряда для функции e^x. Поэтому все сказанное об убывании членов с высокими степенями x в формуле (2) относится и к рядам (4) и (5) — и эти ряды, подобно (2), позволяют вычислить значение соответствующей функции при любом x.

Подчеркнем, однако, сразу же, что убывание членов ряда при любом x, а тем более стремление при любом x «частичных» (конечных) сумм ряда к определенному числу ни в коей мере не являются общим законом: ряды (1), (1а), (2), (4) и (5) являются в определенном смысле более благополучными, чем «общие» ряды (1. 18) и (1. 19). Этому вопросу будет посвящен следующий параграф.

Ряды Тейлора и Маклорена находят множество приложений в высшей математике;

Таблица 2

x	$\varphi°$	$\sin x$	x	$x-\frac{x^3}{6}$	$x-\frac{x^3}{6}+\frac{x^5}{120}$
0	0	0,0000	0,0000	0,0000	0,0000
$\pi/20$	9	0,1564	0,1571	0,1564	0,1564
$\pi/10$	18	0,3090	0,3142	0,3090	0,3090
$3\pi/20$	27	0,4540	0,4712	0,4538	0,4540
$4\pi/20$	36	0,5878	0,6283	0,5869	0,5878
$5\pi/20$	45	0,7071	0,7854	0,7046	0,7071
$6\pi/20$	54	0,8090	0,9425	0,8029	0,8091
$7\pi/20$	63	0,8910	1,0996	0,8780	0,8914
$8\pi/20$	72	0,9510	1,2566	0,9258	0,9519
$9\pi/20$	81	0,9877	1,4137	0,9427	0,9898
$\pi/2$	90	1,0000	1,5708	0,9248	1,0045

Таблица 3

x	$\varphi°$	$\cos x$	$1-\frac{x^2}{2}$	$1-\frac{x^2}{2}+\frac{x^4}{24}$	$1-\frac{x^2}{2}+\frac{x^4}{24}-\frac{x^6}{720}$
0	0	1,0000	1,0000	1,0000	1,0000
$\pi/20$	9	0,9877	0,9877	0,9877	0,9877
$\pi/10$	18	0,9510	0,9506	0,9510	0,9510
$3\pi/20$	27	0,8910	0,8890	0,8911	0,8910
$4\pi/20$	36	0,8090	0,8026	0,8091	0,8090
$5\pi/20$	45	0,7071	0,6916	0,7075	0,7071
$6\pi/20$	54	0,5878	0,5558	0,5887	0,5877
$7\pi/20$	63	0,4540	0,3954	0,4563	0,4539
$8\pi/20$	72	0,3090	0,2105	0,3144	0,3089
$9\pi/20$	81	0,1564	0,0007	0,1672	0,1561
$\pi/2$	90	0,0000	—2,337	0,0200	—0,0009

с некоторыми из них мы встретимся ниже (см., например, § 6). Здесь же укажем на возможность использования рядов для и н т е г р и р о в а н и я функций. Если функция $y=f(x)$ раскладывается в степенной ряд

$$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n+\ldots, \qquad (6.2.6)$$

то, вообще говоря (с некоторыми оговорками, связанными с содержанием § 3; здесь мы пока этих ограничений не коснемся),

$$\int f(x)\,dx=\int (a_0+a_1x+a_2x^2+\ldots$$
$$\ldots+a_nx^n+\ldots)\,dx=C+a_0x+\frac{a_1}{2}x^2+$$
$$+\frac{a_2}{3}x^3+\ldots+\frac{a_n}{n+1}x^{n+1}+\ldots \qquad (6.2.7)$$

При этом результат (7) справедлив независимо от того, можно ли выразить неопределенный интеграл (первообразную функцию) $\int f(x)\,dx$ с помощью алгебраической формулы или нельзя. Так, например, выше мы видели, что функция $\int e^{-t^2}\,dt$ не может быть выражена никакой (конечной) формулой. Однако «бесконечной формулой» (рядом) эту функцию представить можно. Подставив в (2) вместо x выражение $-t^2$, получим

$$e^{-t^2}=1-t^2+\frac{t^4}{2}-\frac{t^6}{6}+\ldots$$
$$\ldots+(-1)^n\frac{t^{2n}}{n!}+\ldots; \qquad (6.2.8)$$

поэтому

$$\int e^{-t^2}dt=C+t-\frac{t^3}{3}+\frac{t^5}{10}-$$
$$-\frac{t^7}{42}+\ldots+(-1)^n\frac{t^{2n+1}}{n!\,(2n+1)}+\ldots \qquad (6.2.9)$$

В частности[1],

$$\int\limits_0^x e^{-t^2}dt=x-\frac{x^3}{3}+\frac{x^5}{2\cdot 5}-\frac{x^7}{6\cdot 7}+\ldots+$$
$$+(-1)^n\frac{x^{2n+1}}{n!\,(2n+1)}+\ldots \qquad (6.2.10)$$

---

[1] Формула (10) применима при любом к о н е ч н о м x; однако ясно, что важный результат $\int\limits_0^\infty e^{-t^2}dt=\frac{\sqrt{\pi}}{2}$ вывести из (10) нельзя — здесь нужно использовать совсем другие приемы.

Упражнения

6.2.1. Выразите коэффициенты c_0, c_1, c_2, ... «ряда» (многочлена) (1а) через коэффициенты b_0, b_1, b_2, ... многочлена (1).

6.2.2. Известно, что функция

$$\operatorname{si} x = \int \frac{\sin x}{x}\, dx \quad (\textit{интегральный синус})$$

не выражается явной формулой. Представьте функцию $\operatorname{si} x$ в виде бесконечного ряда.

§ 3. Случаи неприменимости рядов. Геометрическая прогрессия

Формулы, в которых функция представляется как сумма ряда из степеней x с постоянными коэффициентами, были составлены в предыдущем параграфе для четырех функций: многочлена, e^x, $\sin x$, $\cos x$. Во всех этих случаях оказалось, что при любом x каждый следующий член ряда, исключая, может быть, несколько первых членов, меньше предыдущего и чем больше номер члена, тем ближе этот член к нулю (а в первом — самом благополучном — случае вообще все члены ряда, начиная с некоторого, обращаются в нуль). В этих примерах при любом x можно вычислить значение функции при помощи ряда, если взять достаточное число членов ряда так, чтобы отброшенные члены практически не влияли на результат.

Напомним, что мы начинали с задачи о приближенном выражении функции в малой области изменения переменной и строили все более точные формулы с учетом первой, второй, третьей и т. д. производных. Точность каждой формулы

$$y(x) \approx y(a), \tag{0}$$

$$y(x) \approx y(a) + (x-a)\, y'(a), \tag{I}$$

$$y(x) \approx y(a) + (x-a)\, y'(a) + \frac{(x-a)^2}{2} y''(a) \tag{II}$$

и т. д. тем больше, чем меньше величина $|x-a|$. Кроме того, при данном $x-a$ каждое последующее приближение точнее предыдущего: формула (I) точнее формулы (0), формула (II) точнее формулы (I) и т. д. Поэтому увеличение числа членов ряда позволяет увеличить величину $|x-a|$, сохраняя заданную точность.

Теперь поставим вопрос, всегда ли можно при **любом** значении $|x-a|$ достичь заданной точности путем увеличения числа членов ряда. На очень важном примере мы убедимся сейчас, что это не так. Степенной ряд, дающий хорошее приближение в малой области изменения x, т. е. при малых $|x-a|$, может быть принципиально непригодным в области больших $|x-a|$. Другими словами, этот ряд может иметь естественный предел применимости, предел допустимого увеличения $|x-a|$, не зависящий от числа взятых членов ряда. Для примеров предыдущего параграфа дело обстояло более благополучно, но столь хорошее положение дел является исключением, а не общим правилом.

Рассмотрим следующую простую функцию:

$$y = \frac{1}{1-x} = (1-x)^{-1}.$$

Вычисляя последовательно производные, найдем

$$y' = \frac{1}{(1-x)^2}, \quad y'' = \frac{1\cdot 2}{(1-x)^3}, \ \ldots, \quad y^{(n)} = \frac{n!}{(1-x)^{n+1}}, \ \ldots$$

Для $x=0$ будем иметь

$$y(0) = 1, \quad y'(0) = 1, \quad y''(0) = 2, \ldots \ldots, y^{(n)}(0) = n!, \ \ldots$$

Таким образом, получим ряд

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots \ldots + x^n + \ldots \tag{6.3.1}$$

Пример функции $\frac{1}{1-x}$ замечателен не только необычайно простым видом получившегося степенного ряда (все коэффициенты равны 1). В этом случае нетрудно дать точную формулу для суммы n первых членов ряда (1):

$$1 + x + x^2 + \ldots + x^{n-1} = \frac{1-x^n}{1-x}. \tag{6.3.2}$$

Справедливость этой формулы — формулы суммы конечной *геометрической прогрессии* — должна быть известна из курса средней школы. (Ее можно доказать, умножая обе части формулы (2)

на $1-x$.) Соотношение (2) можно переписать так:

$$1+x+x^2+\ldots+x^{n-1}=$$
$$=\frac{1}{1-x}-\frac{x^n}{1-x}. \qquad (6.3.3)$$

Сравнивая (3) и (1), видим, что $\frac{x^n}{1-x}$ есть «остаточный член» приближенного равенства

$$\frac{1}{1-x}\approx 1+x+x^2+\ldots+x^{n-1}, \qquad (6.3.4)$$

т. е. та величина, которой мы пренебрегаем, если ограничиваемся первыми n членами стоящего в правой части (1) ряда

$$1+x+x^2+x^3+\ldots+x^n+\ldots \qquad (6.3.1a)$$

Если $-1<x<1$, то x^n тем ближе к нулю, чем больше n, и, следовательно, взяв достаточно много членов ряда, мы отбрасываем малую величину. Ряды, обладающие таким свойством, называются *сходящимися*. Сходимость ряда (1а) при $x=\frac{1}{2}$ иллюстрирует приведенная ниже таблица; здесь $\frac{1}{1-x}=2$. Величина Δ в таблице — это выраженная в процентах погрешность приближенной формулы (4) (разность левой [redacted] правой частей (4) в отношении к левой части).

n	1	2	3
$1+x+x^2+\ldots+x^{n-1}$	1	1,5	1,75
Δ	50%	25%	12,5%

n	4	5	6
$1+x+x^2+\ldots+x^{n-1}$	1,875	1,9325	1,9637
Δ	6,25%	3,12%	1,56%

Заметим, что чем ближе x к 1, тем больше членов ряда приходится брать для получения заданной точности (тем хуже сходится ряд, как говорят математики).

Однако картина полностью изменится, если взять $|x|\geqslant 1$. Так, если $x>1$, то каждый следующий член ряда (1а) больше предыдущего. Формула (3) остается в силе, однако при $x>1$ величина x^n неограниченно растет вместе с ростом n, [redacted] поэтому дробью $\frac{[redacted]^n}{1-x}$ пренебрегать никак нельзя. Формула (1) в этом случае будет уже неверна. Нет даже никакого качественного сходства между суммой положительных слагаемых (1а) и отрицательной (при $x>1$) величиной $\frac{1}{1-x}$. Так, при $x=2$ формула (1) обращается в абсурдное «равенство»

$$-1=1+2+4+8+16+\ldots$$

При $x>1$ сумма ряда (1а) неограниченно увеличивается при увеличении n (это видно и из (3)). Если же $x\leqslant -1$, то мы получаем в правой части (1) знакопеременную сумму, никак не похожую на величину $\frac{1}{1-x}$. Так, при $x=-1$ формула (1) дает

$$1/2=1-1+1-1+1-\ldots$$

Ряды, подобные тем, в которые обращается (1а) при $|x|\geqslant 1$, называются *расходящимися*.

Мы установили, что сумма членов бесконечной геометрической прогрессии (1а) равна $\frac{1}{1-x}$, если $|x|<1$; если же $|x|\geqslant 1$, то бесконечная геометрическая прогрессия не имеет конечной суммы (расходится), другими словами, сумма первых n членов этого ряда при $n\to\infty$ не стремится ни к какому пределу.

Отметим еще, что любая периодическая (десятичная) дробь представляет собой сумму членов геометрической прогрессии. Так, например,

$$1,(1)=1,1111\ldots=1+0,1+0,01+$$
$$+0,001+\ldots=1+0,1+(0,1)^2+$$
$$+(0,1)^3+\ldots=\frac{1}{1-0,1}=\frac{1}{0,9}=$$
$$=\frac{10}{9}=1\frac{1}{9}.$$

Таким образом, с простейшим рядом (геометрической прогрессией) мы встречались уже [redacted] раньше, в арифметике и алгебре.

Функция $y=\frac{1}{1-x}$ (рис. 1) терпит разрыв при $x=1$: если x близок к 1, но больше 1, то $\frac{1}{1-x}$ — большое по аб-

солютной величине отрицательное число; если x близок к 1, но меньше 1, то $\frac{1}{1-x}$ — большое положительное число. Следовательно, при переходе x через значение $x=1$ функция $\frac{1}{1-x}$ скачком переходит от больших положительных чисел к большим по абсолютной величине отрицательным числам. Этой особенности поведения функции ряд описать не может, поэтому он и отказывается нам служить.

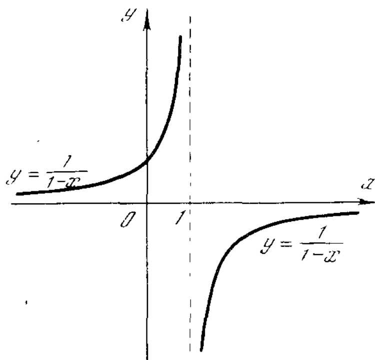


Рис. 6.3.1

Отметим еще следующее: при $x=1$ функция $y=\frac{1}{1-x}$ обращается в бесконечность (чем ближе x к 1, тем больше y по абсолютной величине), и при этом же значении x члены ряда (1а) перестают убывать. Ряд может быть пригоден для вычислений, лишь если его члены достаточно быстро стремятся к нулю, т. е. убывают по абсолютной величине [1]. При $x=1$ формула (1) неверна, так как члены стоящего в правой части ряда не убывают. Значит, ряд (1а) непригоден для вычисления значений функции $y=\frac{1}{1-x}$ и при $x=-1$, и при $x<-1$ (здесь тоже члены (1а) не убывают), хотя сама функция y при $x=-1$ разрыва не испытывает и равна $\frac{1}{1-(-1)}=\frac{1}{2}$.

Как бы мы ни выбирали коэффициенты многочлена, его график всегда будет сплошной, непрерывной линией: график многочлена разрывов не имеет. Поэтому если некоторая функция $f(x)$ терпит разрыв при $x=x_0$ (в примере с $\frac{1}{1-x}$ мы имели $x_0=1$), то при значении $x=x_0$ ряд, составленный для $f(x)$, заведомо непригоден для вычислений. Поскольку каждый член c_nx^n ряда $c_0+c_1x+c_2x^2+\ldots$ будет тем больше по абсолютной величине, чем больше абсолютная величина x, то при любом x, по абсолютной величине большем, чем x_0, ряд также непригоден для вычислений.

Таким образом, при наличии разрыва $f(x)$ можно заранее указать такое x_0, что при всех x, больших x_0 по абсолютной величине, ряд будет непригоден для вычислений.

Рассмотрим еще один пример. Составим ряд Маклорена для функции $y=\operatorname{tg} x$. По общим правилам находим:

$$y=\operatorname{tg} x=\frac{\sin x}{\cos x}, \quad y'=\frac{1}{\cos^2 x},$$

$$y''(x)=\frac{2\sin x}{\cos^3 x}, \quad y'''(x)=\frac{2+4\sin^2 x}{\cos^4 x},$$

$$y^{\mathrm{IV}}(x)=\frac{16\sin x+8\sin^3 x}{\cos^5 x},$$

$$y^{\mathrm{V}}(x)=\frac{16+88\sin^2 x+16\sin^4 x}{\cos^6 x}, \ldots$$

Отсюда $y(0)=0$, $y'(0)=1$, $y''(0)=0$, $y'''(0)=2$, $y^{\mathrm{IV}}(0)=0$, $y^{\mathrm{V}}(0)=16$, ... Поэтому

$$\operatorname{tg} x=0+1\cdot x+0\cdot x^2+ + \frac{2}{3\cdot 2\cdot 1}x^3+0\cdot x^4+ + \frac{16}{5\cdot 4\cdot 3\cdot 2\cdot 1}x^5+\ldots$$

Таким образом,

$$\operatorname{tg} x=x+\frac{1}{3}x^3+\frac{2}{15}x^5+ + \frac{17}{315}x^7+\frac{62}{2835}x^9+\ldots \qquad (6.3.5)$$

(здесь коэффициенты при x^7 и x^9 могут быть получены так же, как были получены коэффициенты при x, x^3 и x^5).

Что можно сказать об области применимости ряда (5)? График тангенса (см. рис. 4.10.5) подсказывает нам, что ряд (5) может быть пригоден для вычислений лишь при $|x|<\frac{\pi}{2}$, поскольку

[1] Конечно, если один-два или несколько первых членов ряда возрастают, это не беда — при условии, что следующие дальше члены ряда достаточно быстро убывают (см. пример с e^x при $x=2$ (табл. 1 из § 2)).

при $x=\frac{\pi}{2}$ функция $\operatorname{tg} x$ ведет себя так же плохо, как функция $\frac{1}{1-x}$ при $x=1$.

Однако по форме ряда (5) нелегко сказать, при каком значении x этот ряд нельзя применять, ибо закон образования коэффициентов ряда в отличие от рассмотренного выше ряда (1а) вовсе не прост [2].

Отметим, что наличие разрыва функции достаточно для того, чтобы ряд перестал сходиться, но не необходимо. В качестве примера рассмотрим функцию $y=\frac{1}{1+x}$. Применяя формулу (1.19) или просто заменяя в равенстве (1) x на $(-x)$, найдем

$$\frac{1}{1+x}=1-x+x^2-x^3+\ldots \qquad (6.3.6)$$

Возьмем, например, $x=2$. Тогда

$$\left.\frac{1}{1+x}\right|_{x=2}=\frac{1}{1+2}=\frac{1}{3}.$$

Сумма же членов ряда

$$1-x+x^2-x^3+\ldots \qquad (6.3.7)$$

при изменении числа n первых его членов колеблется весьма резко и ни при каком n на $1/3$ не похожа:

n	1	2	3	4	5	6	7	...
Сумма n членов ряда (7)	1	−1	3	−5	11	−21	43	...

Ясно, что при $x=2$ наш ряд непригоден для вычисления значения функции $y=\frac{1}{1+x}$. Но почему дело обстоит таким образом? Ведь сама функция y нигде в интервале от $x=0$ до $x=2$ никаких разрывов не имеет; в этих пределах она — гладкая, «хорошая» функция (рис. 2).

Отказ ряда (7) «работать» при $x=2$ объясняется тем, что функция $y=\frac{1}{1+x}$ испытывает разрыв при $x=-1$. Поэтому при $x=-1$ абсолютные величины членов ряда (7) не убывают с номером рассматриваемого члена. Но абсолютные величины членов ряда (7) не зависят от знака x. Поэтому и при $x=1$, и тем более при $x>1$ этот ряд для вычислений непригоден.

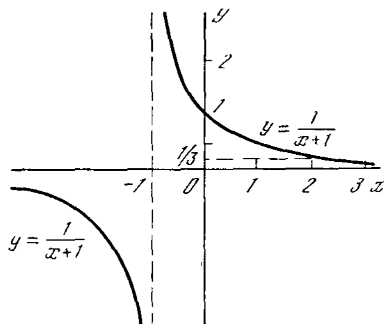


Рис. 6.3.2

Таким образом, даже если нас интересует лишь поведение ряда при положительных x, все равно надо учитывать все значения x, в том числе и отрицательные, при которых разлагаемая в ряд функция претерпевает разрыв.

Для читателей, знакомых с *комплексными числами* (см. главы 14 и 15), укажем, что

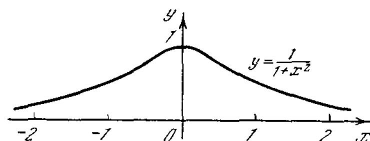


Рис. 6.3.3

на самом деле на сходимость ряда при вещественных значениях x влияет даже поведение рассматриваемой функции при комплексных значениях аргумента. Приведем пример этого (см. также § 15.2). Заменив в формуле (6) x на x^2, получим

$$\frac{1}{1+x^2}=1-x^2+x^4-x^6+\ldots \qquad (6.3.8)$$

График функции $y=\frac{1}{1+x^2}$ (рис. 3) не имеет разрывов ни при положительных, ни при отрицательных x, нигде не уходит в бесконечность. Однако ряд (8) пригоден для вычислений, лишь если $x^2<1$, т. е. при $-1<x<1$. Причина этого в том, что при $x=\pm\sqrt{-1}=\pm i$,

[2] Заметим, что отношения последовательных коэффициентов ряда (5) $1:1/3=3$; $1/3:2/15=2{,}5$; $2/15:17/315\approx 2{,}4706$; $17/315:62/2835\approx$ $\approx 2{,}4677$ быстро сходятся к величине $\left(\frac{\pi}{2}\right)^2\approx$ $\approx 2{,}4674$. Мы, однако, никак не можем остановиться здесь на доказательстве этого неожиданного факта.

т. е. при $x^2=-1$, функция $y=\frac{1}{1+x^2}$ обращается в бесконечность; поэтому члены ряда не убывают по абсолютной величине при $x^2=-1$, а значит, и при $x^2=1$ (ср. гл. 15, с. 430).

Можно привести и другие примеры функций, которые нельзя разложить в ряд Тейлора, например функцию e^{-1/x^2} при $x=0$ (ср. с. 169). Не разлагаются в ряд Маклорена дробные степени x и функции, содержащие дробную степень аргумента, например функция $\sin\sqrt{x}$.

В рамках данной книги разъяснить достаточно подробно и понятно вопрос о границах применимости формул Тейлора и Маклорена невозможно; для того чтобы разобраться в возникающих здесь вопросах, надо обратиться к какому-либо курсу теории функций комплексного переменного (ср. гл. 15). Остается одно утешение — все встречающиеся физику или инженеру функции не разлагаются в ряд Тейлора только в отдельных (особых) точках. Как правило, мы встречаемся с функциями, которые можно, во всяком случае на удачно подобранном конечном интервале, разложить в ряд Тейлора.

Упражнения

6.3.1. Выпишите ряд Маклорена для функций: а) $y=\frac{x+1}{1-x}$; б) $y=\ln(1+x)$.

6.3.2. Выпишите разложение функции $y=\ln x$ в ряд Тейлора по степеням $x-1$.

Каковы области применимости рядов, полученных в упр. 1, 2?

6.3.3. Пусть $f(x)$ и $g(x)$ — известные функции. Найдите первые три члена разложения в ряд по степеням x произведения функций $f(x)\,g(x)$. Постройте тот же ряд перемножением ряда для $f(x)$ и ряда для $g(x)$; сравните полученные результаты.

§ 4. Бином Ньютона для целых и дробных показателей

Составим разложение в ряд Маклорена произвольной степени m двучлена $(a+x)$, т. е. функции $y=(a+x)^m$.

По общему правилу сначала найдем производные

$$y'=m(a+x)^{m-1},$$

$$y''=m(m-1)(a+x)^{m-2},\ \ldots,$$

$$y^{n}=m(m-1)\ldots(m-n+1)\times(a+x)^{m-n},\ \ldots \tag{6.4.1}$$

и значения самой функции и ее производных при $x=0$:

$$y(0)=a^m,\quad y'(0)=ma^{m-1},$$

$$y''(0)=m(m-1)a^{m-2},\ \ldots, \tag{6.4.2}$$

$$y^{n}(0)=m(m-1)\ldots(m-n+1)a^{m-n},\ \ldots$$

Отсюда получим ряд Маклорена

$$(a+x)^m=a^m+\frac{m}{1}a^{m-1}x+\frac{m(m-1)}{1\cdot 2}\times a^{m-2}x^2+\ldots$$

$$\ldots+\frac{m(m-1)(m-2)\ldots(m-n+1)}{n!}\times a^{m-n}x^n+\ldots \tag{6.4.3}$$

Если показатель степени m — положительное целое число, то $(a+x)^m$ есть многочлен степени m, так что в этом случае ряд (3) будет конечным: производная $(m+1)$-го порядка функции $(a+x)^m$ равна нулю, а значит, и все ее производные более высоких порядков тоже равны нулю. Формулы (1)—(3) отражают это обстоятельство. В самом деле, при $n=m+1$ множитель $(m-n+1)$ обращается в нуль; при $n>m+1$ где-то в последовательности множителей m, $(m-1)$,. . . найдется множитель, равный нулю, и, следовательно, произведение будет равно нулю.

При целом положительном m произведение в числителе можно записать в более удобной форме: умножив и разделив произведение $m(m-1)\ldots(m-n+1)$ на $(m-n)(m-n-1)\ldots 3\cdot 2\cdot 1$, получим

$$m(m-1)\ldots(m-n+1)=\frac{m(m-1)\ldots 3\cdot 2\cdot 1}{(m-n)(m-n-1)\ldots 3\cdot 2\cdot 1}=\frac{m!}{(m-n)!}.$$

Окончательно будем иметь

$$(a+x)^m=a^m+\frac{m!}{1!(m-1)!}a^{m-1}x+\frac{m!}{2!(m-2)!}a^{m-2}x^2+\ldots$$

$$\ldots+\frac{m!}{n!(m-n)!}a^{m-n}x^n+\ldots$$

$$\ldots+\frac{m!}{(m-2)!2!}a^2x^{m-2}+\frac{m!}{(m-1)!1!}ax^{m-1}+x^m. \tag{6.4.4}$$

В формуле (4) справа и слева стоят многочлены степени m. Таким образом, для целого положительного m мы имеем точное равенство (4), справедливое при любых значениях x. Правая часть (4) симметрична относительно x и a: коэффициенты при членах $a^{m-n}x^n$ и $a^n x^{m-n}$ равны. Это понятно, так как $(x+a)^m$ не зависит от порядка слагаемых в скобке: ясно, что $(x+a)^m = (a+x)^m$.

Формулу (4) не совсем справедливо (см. ниже) называют формулой *бинома Ньютона*. Ее можно получить, не пользуясь производными и рядами. Для этого нужно взять произведение m сомножителей $(a+x)(a+x)\cdots(a+x)$, выполнить умножение и привести подобные члены. Однако при m, заданном в общем виде буквой, а не числом, приведение подобных членов становится довольно трудным, и в целом вывод бинома Ньютона при помощи высшей математики, т. е. со ссылкой на ряд Маклорена, оказывается проще элементарного вывода той же формулы.

Отметим, что Ньютон получил общую формулу (3), т. е. разложение для $(x+a)^m$, в случае л ю б ы х показателей m. Поэтому правильнее было бы именно формулу (3) называть формулой бинома Ньютона, а не формулу (4), которая была известна и до Ньютона [1] (и которая представляет собой простой частный случай зависимости, связывающей коэффициенты представлений (2.1) и (2.1а) многочлена; см. упр. 2.1).

Вернемся к общей формуле (3). Если m не целое положительное, то, поскольку степени n переменной x — целые положительные, ни один из множителей $m, m-1, \ldots, m-n+1$ числителя коэффициента при x^n не обращается в нуль. Таким образом, все эти коэффициенты отличны от нуля и формула (3) дает б е с к о н е ч н ы й ряд. В частности, при $m=-1$ этот ряд принимает вид

$$\frac{1}{a+x}=\frac{1}{a}-\frac{x}{a^2}+\frac{x^2}{a^3}-\frac{x^3}{a^4}+\ldots \qquad (6.4.5)$$

Заметим, что при $a=1$ формула (5) переходит в уже знакомую нам формулу суммы членов (бесконечной) геометрической прогрессии (3.6).

Из формулы (5) находим также

$$\frac{1}{a-x}=\frac{1}{a}+\frac{x}{a^2}+\frac{x^2}{a^3}+\frac{x^3}{a^4}+\ldots$$

Для $m=\frac{1}{2}$ получим

$$\sqrt{a+x}=\sqrt{a}+\frac{1}{2}\frac{x}{\sqrt{a}}-\frac{1}{8}\frac{x^2}{a\sqrt{a}}+\frac{1}{16}\frac{x^3}{a^2\sqrt{a}}-\frac{5}{128}\frac{x^4}{a^3\sqrt{a}}+\frac{7}{256}\frac{x^5}{a^4\sqrt{a}}-\frac{21}{1024}\frac{x^6}{a^5\sqrt{a}}+\ldots \qquad (6.4.6)$$

В разложении $(a+x)^m$ при любом m все члены имеют одинаковую сумму степеней при a и при x и каждый следующий член отличается от предыдущего множителем $\left(\frac{x}{a}\right)$ и коэффициентом. Физик сказал бы, что a и x в формуле (3) должны быть одинаковой размерности, значит, отношение $\frac{x}{a}$ безразмерно. Можно с самого начала вынести за скобку a

$$(a+x)^m=a^m\left(1+\frac{x}{a}\right)^m$$

и разлагать $\left(1+\frac{x}{a}\right)^m$ по степеням $\frac{x}{a}$.

Оказывается, что при всех m (дробных, положительных и отрицательных) ряд (3) пригоден (т. е. сходится) лишь при $\left|\frac{x}{a}\right|<1$, т. е. при $|x|<|a|$. При $\left|\frac{x}{a}\right|>1$ ряд (3) — расходящийся. Исключение представляют целые положительные m, потому что в этом случае формула (3) содержит конечное число членов.

Принимая в (3) $a=1$ и заменив букву x на r (где должно быть $|r|<1$), мы, ограничившись лишь первыми двумя слагаемыми правой части (3), получим приближенную формулу

$$(1+r)^m \approx 1+mr,$$

которой неоднократно пользовались выше; она справедлива лишь при малых $|r|$ (более точная формула имеет вид $(1+r)^m \approx 1+mr+\frac{m(m-1)}{2}r^2$).

Формула (6) дает хороший способ вычисления квадратных корней. При этом,

[1] В Европе формулу (4) (где m — натуральное число) впервые получил итальянский математик Никколо Т а р т а л ь я (ок. 1499—1557); однако еще раньше она была известна арабским математикам.

чем меньше $\left|\frac{x}{a}\right|$, тем меньше членов можно брать в (6) для достижения заданной точности.

Упражнения

6.4.1. Используя разложение в ряд, найдите $\sqrt{1,1}$ и $\sqrt{1,5}$ как $\sqrt{1+x}$ при $x=0,1$ и при $x=0,5$, удерживая в разложении два, три и четыре члена. Сравните полученные результаты с табличными значениями.

6.4.2. Покажите, что при $|x|<1$: а) $\sqrt[n]{1+x}\approx 1+\frac{x}{n}$; б) $\sqrt[n]{1+x}\approx 1+\frac{x}{n}-\frac{n-1}{2n^2}x^2$, причем формулы а) и б) тем точнее, чем меньше $|x|$ (и формула б) точнее формулы а)).

6.4.3. Найдите по формулам упр. 2 $\sqrt[3]{1,2}$, $\sqrt[3]{1,1}$, $\sqrt[3]{1,05}$. Сравните полученные значения с найденными из таблиц.

6.4.4. Найдите $\sqrt{6}$ с тремя верными знаками после запятой. [Указание. Воспользуйтесь тем, что $6=4+2$, $\sqrt{4}=2$ и примените формулу (6).]

6.4.5. Почему нельзя разложить $y=\sqrt{x}$ по формуле Маклорена? А можно ли разложить по этой формуле функцию $y=\sqrt[3]{x^{1000}}$?

§ 5. Порядок возрастания и убывания функций. Правило Бернулли—Лопиталя

Разложение функций в ряд дает общий способ приведения различных функций к однотипному виду и позволяет сравнивать между собой различные функции. Функции надо сравнивать, например, в тех случаях, когда рассматривается отношение $\frac{f(x)}{g(x)}$ при таком значении аргумента x, при котором как $f(x)$, так и $g(x)$ близки к нулю. Такие отношения часто можно приравнять определенному числу — не очень малому, т. е. не близкому к нулю, но и не очень большому; в этом нас убеждают примеры на вычисление производных. Однако иногда отношение двух близких к нулю функций может обращаться в нуль или в (положительную или отрицательную) бесконечность, и эти случаи надо уметь распознавать. Приведем несколько примеров. Для простоты рассмотрим случаи, в которых интересующее нас значение x равно нулю.

При малом x функции $\sin x$ и $\operatorname{tg} x$ также малы. Функции e^x и $\cos x$ близки к 1, а, следовательно, e^x-1 и $1-\cos x$ малы. При этом значения функций $\sin x$, $\operatorname{tg} x$, e^x-1 и $1-\cos x$ будут тем ближе к нулю, чем меньше $|x|$.

Сравним эти функции с величиной x. Для этого разложим функции в ряды Маклорена:

$$\begin{aligned} \sin x&=x-\frac{x^3}{6}+\ldots,\\ \operatorname{tg} x&=x+\frac{x^3}{3}+\ldots,\\ 1-\cos x&=\frac{x^2}{2}-\frac{x^4}{24}+\ldots,\\ e^x-1&=x+\frac{x^2}{2}+\ldots \end{aligned} \tag{6.5.1}$$

Отсюда находим

$$\frac{\sin x}{x}=1-\frac{x^2}{6}+\ldots$$

Следовательно,

$$\frac{\sin x}{x}\to 1 \text{ при } x\to 0, \text{ или } \lim_{x\to 0}\frac{\sin x}{x}=1.$$

Аналогично из (1) находим:

$$\frac{\operatorname{tg} x}{x}=1+\frac{x^2}{3}+\ldots\to 1 \text{ при } x\to 0;$$

$$\frac{1-\cos x}{x^2}=\frac{1}{2}-\frac{x^2}{24}+\ldots\to\frac{1}{2} \text{ при } x\to 0;$$

$$\frac{1-\cos x}{x}=\frac{x}{2}-\frac{x^3}{24}+\ldots\to 0 \text{ при } x\to 0;$$

$$\frac{e^x-1}{x}=1+\frac{x}{2}+\ldots\to 1 \text{ при } x\to 0.$$

Можно найти и более сложные соотношения. Например, из того, что

$$\sin x=x-\frac{x^3}{6}+\frac{x^5}{120}-\ldots,$$

$$\operatorname{tg} x=x+\frac{x^3}{3}+\frac{2}{15}x^5+\ldots,$$

следует

$$\operatorname{tg} x-\sin x=\frac{1}{2}x^3+\frac{1}{8}x^5+\ldots,$$

и, значит,

$$\frac{\operatorname{tg} x-\sin x}{x^3}\to\frac{1}{2} \text{ при } x\to 0.$$

Можно построить шкалу порядка убывания различных функций при стремлении x к нулю. Если $f(x)\to 0$ при $x\to 0$, то *порядком убывания* (или *порядком малости*) $f(x)$ относительно x назовем степень x, убывающую так же быстро, как и $f(x)$. Точнее, утвержде-

ние о том, что $f(x)$ имеет порядок малости k относительно x, означает, что эта функция убывает, как x^k, т. е. что отношение $\frac{f(x)}{x^k}$ имеет пределом при $x \to 0$ конечное, не равное нулю число.

Таким образом, $\sin x$, $\operatorname{tg} x$, $e^x - 1$ при $x \to 0$ убывают, имея первый порядок малости, $1-\cos x$ имеет второй порядок малости, $\operatorname{tg} x - \sin x$ — третий порядок малости относительно x.

В некоторых конкретных случаях порядок малости можно определить и без разложения в ряд. Например, из широко известного рис. 1, где MP и NA — линии синуса и тангенса, отвечающие $\angle AOM = x$, легко следует, что $\operatorname{tg} x > x > \sin x$ (соответствующее рассуждение, основанное на том, что $S_{\Delta OMA} < S_{\text{сект.} OMA} < S_{\Delta ONA}$, имеется в школьных учебниках). А так как $\sin x \approx \operatorname{tg} x$ при малых x (ибо в этом случае отношение $\frac{\sin x}{\operatorname{tg} x} = \cos x \approx 1$), то $\sin x \approx x$ и $\operatorname{tg} x \approx x$, когда x мало̀, т. е. $\sin x$ и $\operatorname{tg} x$ имеют первый порядок малости. Справедлива формула $1-\cos x = 2\sin^2(x/2)$, а так как $\sin(x/2)$ первого порядка малости, то ясно, что функция $1-\cos x$ должна иметь второй порядок малости. Функцию $\operatorname{tg} x - \sin x$ можно записать как $\frac{\sin x}{\cos x} - \sin x = \frac{\sin x}{\cos x}(1-\cos x)$. Так как для малых x величина $\sin x$ первого порядка, $1-\cos x$ — второго порядка малости, а $\cos x \approx 1$, то ясно, что функция $\operatorname{tg} x - \sin x$ — третьего порядка малости. Однако эти конкретные приемы требуют догадки, остроумия, — и именно поэтому полезен простой безотказный общий метод.

Такое соотношение между остроумным решением отдельных задач и общими методами наблюдается везде: свойства касательных к параболе, площадь круга, объем пирамиды, объем шара известны были древним грекам, но только дифференциальное и интегральное исчисления дали общие простые способы решения **в с е х** задач такого типа, доступные каждому инженеру, физику, учащемуся, в то время как изощренные методы античных мудрецов рассчитаны были разве что на выдающихся ученых.

При помощи рядов можно находить не только отношения функции к степени x, но и отношение одной функции к другой. Приведем примеры этого:

$$\frac{e^x-1}{\sin x} = \frac{x + \frac{x^2}{2} + \frac{x^3}{6} + \dots}{x - \frac{x^3}{6} + \dots} = \frac{1 + \frac{x}{2} + \dots}{1 - \frac{x^2}{6} + \dots} \to 1 \quad \text{при } x \to 0,$$

$$\frac{e^x-1}{1-\cos x} = \frac{x + \frac{x^2}{2} + \dots}{\frac{x^2}{2} - \frac{x^4}{24} + \dots} = \frac{1 + \frac{x}{2} + \dots}{\frac{x}{2} - \frac{x^3}{24} + \dots} \to \infty \quad \text{при } x \to 0,$$

$$\frac{e^x-1}{\sqrt{x}} = \frac{x + \frac{x^2}{2} + \dots}{\sqrt{x}} = \frac{\sqrt{x} + \frac{x^{3/2}}{2} + \dots}{1} \to 0$$

при $x \to 0$.

Коэффициенты рядов Маклорена выражаются через производные. Поэтому

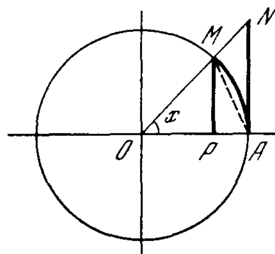


Рис. 6.5.1

можно сформулировать результаты, получаемые при помощи рядов, в виде правил, относящихся к производным. Если $f(0) = g(0) = 0$, то формулы

$$f(x) = f(0) + f'(0)\,x + {}^1/_2 f''(0)\,x^2 + \dots,$$
$$g(x) = g(0) + g'(0)\,x + {}^1/_2 g''(0)\,x^2 + \dots$$

упрощаются так:

$$f(x) = f'(0)\,x + {}^1/_2 f''(0)\,x^2 + \dots,$$
$$g(x) = g'(0)\,x + {}^1/_2 g''(0)\,x^2 + \dots$$

Отсюда

$$\frac{f(x)}{g(x)} = \frac{f'(0)\,x + \frac{1}{2} f''(0)\,x^2 + \dots}{g'(0)\,x + \frac{1}{2} g''(0)\,x^2 + \dots} = \frac{f'(0) + \frac{1}{2} f''(0)\,x + \dots}{g'(0) + \frac{1}{2} g''(0)\,x + \dots} \to \frac{f'(0)}{g'(0)}$$

и, значит,

если $f(0)=g(0)=0$, то $\frac{f(x)}{g(x)} \to \frac{f'(0)}{g'(0)}$ при $x \to 0$ (6.5.2)

(а если также и $f'(0)=g'(0)$, то предел отношения $\frac{f'(x)}{g'(x)}$ при $x \to 0$ равен пределу отношения $\frac{f''(x)}{g''(x)}$ при $x \to 0$[1]).

Разумеется, правило (2) полностью сохраняет силу и в том случае, когда $f(a)=g(a)=0$, где a — какое угодно число, и нас интересует отношение $\frac{f(x)}{g(x)}$ при x, близком к a:

если $f(a)=g(a)=0$, то $\frac{f(x)}{g(x)} \to \frac{f'(a)}{g'(a)}$ при $x \to a$ (6.5.2a)

(а если также и $f'(a)=g'(a)=0$, то пределы отношений $\frac{f(x)}{g(x)}$ и $\frac{f'(x)}{g'(x)}$ при $x \to a$ одинаковы, в силу чего $\frac{f(x)}{g(x)} \to \frac{f''(a)}{g''(a)}$). Здесь для доказательства надо вместо ряда Маклорена использовать ряд Тейлора (см. упр. 2).

Таким образом, если значения двух функций близки к нулю (если обе функции обращаются в нуль при одном и том же значении аргумента, вблизи которого мы эти функции рассматриваем), то отношение самих функций можно заменить отношением их производных. Этот результат называют *правилом Бернулли—Лопиталя*, или (чаще, но менее справедливо) просто *правилом Лопиталя*[2].

После изучения рядов удобнее не запоминать особые правила нахождения пределов, а при малых x пользоваться рядами, в которых функция разложена по степеням x. При этом всюду, где стоит сумма различных степеней x, при переходе к малым x мы оставляем лишь член наименьшей степени.

Аналогично изучению характера убывания функции при малых x (при условии, что при $x=0$ функция обращается в нуль) можно рассматривать поведение функции при б о л ь ш и х x, т. е. при $x \to \infty$. Здесь уместно разлагать функцию по степеням $\frac{1}{x}$ $\Big($ведь $\frac{1}{x}$ у нас мало!$\Big)$, воспользовавшись подстановкой $\frac{1}{x}=t$. Если наша функция является многочленом, то очевидно, что при большом x важен только член с наивысшей степенью x, ибо остальные члены ему уступают: $ax^n+bx^{n-1}+\cdots = ax^n\left(1+\frac{b}{a}\frac{1}{x}+\ldots\right)$; здесь выражение в скобках стремится к 1 при $x \to \infty$. Можно говорить *о порядке роста* функции $f(x)$, т. е. об ее возрастании «как x», «как x^2», «как x^3» и т. д. Говорят, что неограниченно возрастающая по абсолютной величине при $x \to \infty$ функция $f(x)$ имеет порядок роста k, если отношение $\frac{f(x)}{x^k}$ при $x \to \infty$ стремится к конечной (отличной от нуля) величине. При этом ясно, что любой многочлен степени n имеет порядок роста n.

Важнейший факт заключается в том, что при неограниченном росте x *экспоненциальная функция e^x растет быстрее любой степени x* (т. е. e^x имеет, так сказать, «бесконечный порядок роста»). Доказать это легко с помощью разложения (2.2) функции e^x, справедливого, как отмечалось выше, при любом значении x. В силу этого разложения

$$\frac{e^x}{x^n}=\frac{1}{x^n}+\frac{1}{x^{n-1}}+\cdots+\frac{1}{n!}+ + \frac{x}{(n+1)!}+\frac{x^2}{(n+2)!}+\cdots \qquad (6.5.3)$$

При заданном n и при достаточно большом x дробь $\frac{e^x}{x^n}$ станет сколь угодно большой за счет членов в правой части (3) с положительными степенями x. Ясно, что то же относится и к функции e^{kx} с любым положительным k (или к функции $a^x=e^{\ln a \cdot x}$, где $a>1$); так, обозначая $kx=y$, найдем, что

$$\frac{e^{kx}}{x^n}=k^n\frac{e^{kx}}{(kx)^n}=k^n\frac{e^y}{y^n} \to \infty \text{ при } y \to \infty,$$
т. е. при $x \to \infty$. (6.5.4)

Таким образом, любая показательная функция e^{kx}, где $k>0$ (или a^x, где $a>1$), растет быстрее любой степенной функции x^n (или cx^n, где $n>0$ и $c>0$): хотя при небольших x, скажем, график

[1] Так как в силу той же формулы (2) здесь $\lim_{x\to 0}\frac{f'(x)}{g'(x)}=\frac{f''(0)}{g''(0)}$, то мы приходим к равенству $\lim_{x\to 0}\frac{f(x)}{g(x)}=\frac{f''(0)}{g''(0)}$; аналогично, если $f(0)=f'(0)=f''(0)=g(0)=g'(0)=g''(0)=0$, то $\lim_{x\to 0}\frac{f(x)}{g(x)}=\frac{f'''(0)}{g'''(0)}$, и т. д.

[2] Это название сохраняют и за некоторыми родственными сформулированному предложениями, например за тем, о котором говорится в сноске 1 или в упр. 3 и 4.

функции $y=x^{1000}$ уходит вверх гораздо стремительнее графика функции $y_1=e^{0,001x}$ (рис. 2), впоследствии (с ростом x) график функции y_1 снова пересечет график функции y, ибо величина y_1 обгонит величину y.

Показательная функция с отрицательным показателем степени e^{-x}, где $x\to\infty$, *убывает с ростом x быстрее любой отрицательной степени x*: при любом $n>0$

$$f=\frac{e^{-x}}{x^{-n}}=x^n e^{-x}\to 0 \text{ при } x\to\infty. \qquad (6.5.5)$$

Использовать для доказательства (5) разложение e^{-x} по степеням x, считая в нем x очень большим, рискованно, ибо это разложение будет знакопеременным, так что члены разложения в известной степени компенсируют друг друга. Поэтому здесь удобнее рассмотреть обратную f величину

$$\frac{1}{f}=\frac{x^{-n}}{e^{-x}}=\frac{e^x}{x^n}.$$

Согласно (4) при любом n величина $\frac{1}{f}=\frac{e^x}{x^n}\to\infty$ при $x\to\infty$, но это и означает, что $f\to 0$ при $x\to\infty$.

Таким образом, в пределе, при больших по абсолютной величине значениях аргумента, стоящего в показателе, показательная функция (экспонента) зависит от x сильнее, чем любая постоянная степень x: при любом n функция e^x возрастает быстрее x^n, функция e^{-x} убывает быстрее x^{-n}.

Ниже это наглядно продемонстрировано на примере функций x^5 и e^x:

x	1	3	5	10
x^5	1	243	3125	10^5
e^x	2,72	20	150	$2\cdot 10^4$
$\frac{x^5}{e^x}=\frac{e^{-x}}{x^{-5}}$	0,37	12	21	5

x	20	50	100
x^5	$3\cdot 10^6$	$3\cdot 10^8$	10^{10}
e^x	$4\cdot 10^8$	$5\cdot 10^{21}$	10^{43}
$\frac{x^5}{e^x}=\frac{e^{-x}}{x^{-5}}$	0,01	10^{-13}	10^{-33}

Из доказанного вытекает также, что *логарифмическая функция* $\log_a x$ (с любым основанием $a>1$) *растет с ростом x медленнее любой положительной* (хотя бы и очень малой) *степени x^n аргумента x*, а, скажем, функция $e^{-1/x}$ при $x>0$ и $x\to 0$ убывает быстрее любой положительной степени x:

$$\frac{\log_a x}{x^n}\to 0 \text{ при } x\to\infty, \qquad (6.5.6)$$

$$\frac{e^{-1/x}}{x^n}\to 0 \text{ при } x>0 \text{ и } x\to 0. \qquad (6.5.7)$$

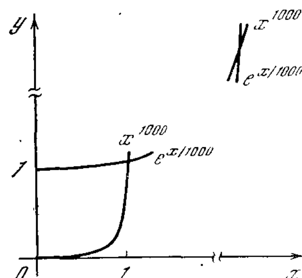


Рис. 6.5.2

Для доказательства (6) достаточно произвести замену $\log_a x=u$, $x=a^u$, $x^n=a^{nu}$, обращающую интересующее нас отношение в $\frac{u}{a^{nu}}=1:\frac{a^{nu}}{u}$, а для доказательства (7) — замену $\frac{1}{x}=v$, обращающую рассматриваемое отношение в $\frac{e^{-v}}{v^{-n}}$ (ср. (5), а также упр. 5).

Также и функция $w=e^{-1/x^2}$ (которую естественно считать равной нулю при $x=0$: ведь при малых по абсолютной величине x, т. е. больших значениях $\frac{1}{x^2}$, и $w=e^{-1/x^2}$ будет сколь угодно мало) убывает при $x\to 0$ быстрее любой степени x, т. е. $\frac{w}{x^n}\to 0$ при всех $n>0$. А отсюда, в свою очередь, вытекает невозможность разложения этой, казалось бы, достаточно «хорошей» функции (см. ее график на рис. 15.1.1) в ряд Маклорена. В самом деле, если бы здесь имела место формула типа (1.19), ряд первых коэффициентов правой части которой может и обратиться в нуль, т. е. формула $w(x)=a_k x^k+a_{k+1}\times\times x^{k+1}+\cdots$, где $a_k\neq 0$, то функция $w(x)$ при малых x имела бы относительно x k-й, а не «бесконечный» порядок малости (ср. с § 15.1).

Упражнения

6.5.1. Найдите следующие пределы: а) $\lim\limits_{x\to 0}\frac{\ln(1+x)}{x}$; б) $\lim\limits_{x\to 0}\frac{\ln(1+x)-x}{x^2}$; в) $\lim\limits_{x\to 0}\frac{\operatorname{tg} x-x}{x^3}$; г) $\lim\limits_{x\to 0}\frac{e^x-1-\operatorname{tg} x}{x^3}$; д) $\lim\limits_{x\to 0}\frac{e^x-1}{\sin x}$; е) $\lim\limits_{x\to 0}\frac{\sin x-x}{x-\operatorname{tg} x}$.

6.5.2. Докажите вариант (2а) правила Бернулли—Лопиталя.

6.5.3. Докажите, что если $f(x)=f'(x)=f''(x)=\ldots=f^{n-1}(x)=0$ при $x=0$ и аналогично $g(0)=g'(0)=g''(0)=\ldots=g^{(n-1)}(0)=0$, то $\lim\limits_{x\to 0}\frac{f(x)}{g(x)}=\frac{f^{(n)}(x)}{g^{(n)}(x)}$ (где мы, разумеется, считаем, что последнее отношение, которое может обращаться в 0 или в ∞, существует).

6.5.4. Докажите, что: а) если $f(x)\to 0$ при $x\to\infty$ и $g(x)\to 0$ при $x\to\infty$, то отношения $\frac{f(x)}{g(x)}$ и $\frac{f'(x)}{g'(x)}$ при $x\to\infty$ стремятся к одному и тому же пределу; б) если $f(x)\to\infty$ при $x\to a$ (где возможен также случай $a=\infty$) и $g(x)\to\infty$ при $x\to a$, то $\lim\limits_{x\to a}\frac{f(x)}{g(x)}=\lim\limits_{x\to a}\frac{f'(x)}{g'(x)}$.

6.5.5. Докажите, что функция $e^{-1/x}$ при $x<0$ и $x\to 0$ растет по абсолютной величине быстрее любой (целой) отрицательной степени x.

§ 6. Дифференциальные уравнения первого порядка. Случай разделяющихся переменных

Выше мы уже встречались (см., например, § 3.6) с понятием **дифференциального уравнения**, связывающего (неизвестную нам) функцию $y=y(x)$ и ее производные:

$$F(x,\ y,\ y',\ y'',\ldots)=0; \tag{6.6.1}$$

здесь F — некоторая известная функция, зависящая от нескольких переменных. Естественнонаучные законы и конкретные свойства тех или иных систем или механизмов весьма часто записываются в виде дифференциальных уравнений, так что во многих случаях существенная часть изучения интересующего нас явления состоит в анализе и решении соответствующего уравнения. В гл. 9 книги сказанное будет проиллюстрировано на примере механики, основу которой составляет так называемый *второй закон Ньютона*, т. е. дифференциальное уравнение (9.4.2).

Высший порядок производной, фигурирующей в дифференциальном уравнении, называется *порядком* уравнения. Здесь мы ограничимся рассмотрением дифференциальных уравнений п е р в о г о порядка:

$$F(x,\ y,\ y')=0, \tag{6.6.2}$$

или, если разрешить такое уравнение относительно производной y':

$$y'=f(x,\ y). \tag{6.6.3}$$

Простейшим случаем уравнения (3) является тот, когда функция f не зависит от y, т. е. когда уравнение имеет вид

$$y'=f(x). \tag{6.6.4}$$

Последнее уравнение указывает, что $y=f(x)$ есть *первообразная* функции $f(x)$, или *неопределенный интеграл*:

$$y=\int f(x)\,dx; \tag{6.6.5}$$

решению уравнения (4) были посвящены главы 3 и 5. Разумеется, уравнение (4) является самым простым, а решение его вытекает из самого определения интеграла; однако, говоря о (общих) дифференциальных уравнениях, полезно иметь в виду этот первый пример, проливающий свет и на общую ситуацию. Так, например, уже здесь хорошо видна неопределенность поставленной задачи: наличие у данного дифференциального уравнения м н о ж е с т в а решений. В случае (4) эти решения имеют вид

$$y=G(x)+C, \tag{6.6.5а}$$

причем $G'(x)=f(x)$. Графики всех функциональных зависимостей $y=y(x)$ получаются один из другого сдвигом (переносом) в направлении оси y (рис. 1). Для того чтобы выделить определенное решение задачи, надо задать значение $y=y_0$ при определенном «начальном» значении $x=x_0$ аргумента — так называемое *начальное условие* для дифференциального уравнения (4); оно определяет единственное решение

$$y=y_0+\int\limits_{x_0}^{x} f(x)\,dx \tag{6.6.5б}$$

уравнения (4) (единственную интегральную кривую $y=y(x)$, проходящую через заданную точку $M_0=M_0(x_0,\ y_0)$).

Теории дифференциальных уравнений посвящено много учебников, содержание которых мы ни в коей мере не намерены здесь повторять. Мы ограничимся здесь лишь тем, что рассмотрим случай уравнений типа (3), которые решаются почти столь же просто (или почти столь же сложно, поскольку они также приводят к вовсе не простой операции интегрирования!), как уравнение (4) для первообразной. А именно, если функция $f(x, y)$ представляет собой произведение $g(x)\,h(y)$ функции от x и функции от y, то переменные x и y в уравнении

$$\frac{dy}{dx}=g(x)\,h(y) \tag{6.6.6}$$

можно р а з д е л и т ь, перенеся в одну часть уравнения все члены, зависящие

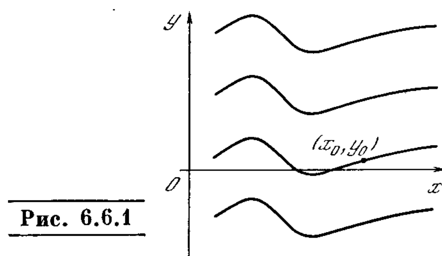


Рис. 6.6.1

от x, а во вторую — члены, зависящие от y. Такие уравнения называются *уравнениями с разделяющимися переменными*.

Разделим обе части (6) на $h(y)$:

$$\frac{1}{h(y)}\,\frac{dy}{dx}=g(x); \tag{6.6.6a}$$

затем умножим обе части (6а) на дифференциал dx:

$$\frac{dy}{h(y)}=g(x)\,dx. \tag{6.6.6б}$$

А теперь для нахождения решения уравнения (6) нам остается лишь взять интегралы от левой и правой частей равенства (6б):

$$\int\frac{dy}{h(y)}=\int g(x)\,dx. \tag{6.6.7}$$

Если стоящие в левой и правой частях (7) интегралы можно вычислить, то мы придем к (неявной) записи функции $y=y(x)$:

$$H(y)=G(x)+C, \tag{6.6.8}$$

где $H(y)=\int\frac{dy}{h(y)}$ и $G(x)=\int g(x)\,dx$ — какие-то значения этих интегралов; C — произвольная постоянная.

Простейшим примером уравнения с разделяющимися переменными может служить уравнение (4), получаемое, если положить в (6) $g\equiv f$ и $h\equiv 1$. Подставив эти значения g и h в (7), мы придем к решению (5) уравнения (4). Почти столь же прост случай, когда в (3) $f(x, y)=f(y)$ зависит только от y, т. е. когда (3) обращается в уравнение

$$y'=f(y); \tag{6.6.4a}$$

здесь, напротив, $g\equiv 1$, а $h\equiv f$. В частности, к этому типу уравнений с разделяющимися переменными относится уравнение

$$y'=ky, \tag{6.6.9}$$

с которым мы в дальнейшем часто будем встречаться (см., например, гл. 8). В случае (9) решение (7) уравнения принимает вид $\int\frac{dy}{y}=k\int dx$, или $\ln y=kx+C$, или, наконец,

$$y=e^{kx+C}=Ae^{kx}, \tag{6.6.10}$$

где обозначено $A=e^C$. (Поскольку C — произвольное число, то и A — произвольное (положительное) число.) Таким образом, мы снова приходим к выводу, что экспоненциальная функция $y=Ae^{kx}$ характеризуется пропорциональностью (9) производной функции и самой функции.

Разберем еще несколько простых примеров уравнений с разделяющимися переменными.

1. Пусть

$$y'=xy, \quad \text{или} \quad \frac{dy}{dx}=xy. \tag{6.6.11}$$

Разделяя переменные, получаем

$$\frac{dy}{y}=x\,dx,$$

откуда $\int\frac{dy}{y}=\int x\,dx$, или $\ln y=\frac{x^2}{2}+C$, или, наконец,

$$y=\exp\left(\frac{x^2}{2}+C\right)=A\exp\left(\frac{x^2}{2}\right), \tag{6.6.12}$$

где по-прежнему обозначено $A=e^C$. Графики функций y изображены на рис. 2,*а*.

2. Пусть

$$y'=\frac{y}{x}, \quad \text{или} \quad \frac{dy}{dx}=\frac{y}{x}; \tag{6.6.13}$$

тогда $\int \frac{dy}{y} = \int \frac{dx}{x}$, или $\ln|y| = \ln|x| + C$, или, наконец,

$$y = Ax \tag{6.6.14}$$

(рис. 2, *б*; здесь снова $|A| = e^C$).

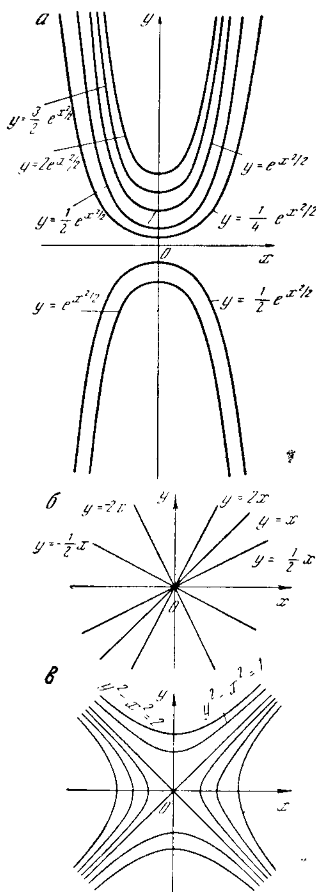


Рис. 6.6.2

3. Если

$$y' = \frac{x}{y}, \quad \text{или} \quad \frac{dy}{dx} = \frac{x}{y}, \tag{6.6.15}$$

то получаем $\int y\,dy = \int x\,dx$, или $\frac{y^2}{2} = \frac{x^2}{2} + C$, или, наконец,

$$y^2 - x^2 = a, \tag{6.6.16}$$

где через a обозначено число $2C$ (рис. 2, *в*).

Разумеется, все эти примеры «удачно подобраны» — здесь не только заданное уравнение (3) есть уравнение вида (6), но и интегралы в решении (7) благополучнейшим образом берутся, так что удается довольно легко выписать общие решения (12), (14) или (16) уравнений. Однако реально дело обстоит столь хорошо вовсе не всегда. В общем случае правая часть уравнения (3), разумеется, вовсе не обязана распадаться на два множителя, зависящие соответственно лишь от x и лишь от y; кроме того, если последнее обстоятельство и имеет место, то интегралы в формуле (7) могут не выражаться элементарными функциями. В таком случае приходится прибегать к тем или иным методам приближенного (численного) решения дифференциальных уравнений.

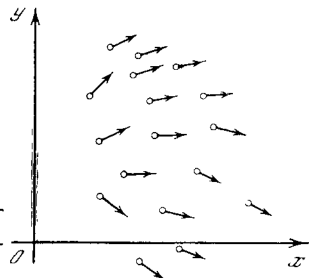


Рис. 6.6.3

Заметим, что уравнение (3) указывает зависимость наклона y' решения $y=y(x)$ этого уравнения от точки (x, y) плоскости; другими словами, уравнение (3) задает *поле направлений* на плоскости, выделяя в каждой точке $M=M(x, y)$ направление с предписанным наклоном $k=y'=f(x, y)$ (рис. 3). Это обстоятельство подсказывает достаточно естественный метод приближенного решения уравнения (3).

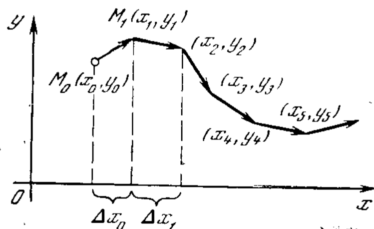


Рис. 6.6.4

Пусть мы хотим найти кривую $y=f(x)$, проходящую через заданную точку $M_0(x_0, y_0)$ плоскости, и такую, что функция $y(x)$ удовлетворяет условию (3). Сместимся из точки M_0 по заданному в ней направлению в близкую точку $M_1(x_1, y_1)$, где $x_1=x_0+\Delta x_0$, $y_1=y_0+\Delta y_0$ и $\frac{\Delta y_0}{\Delta x_0}=k_0=f(x_0, y_0)$. Затем из точки M_1 сместимся по имеющемуся в этой точке направлению в близкую к M_1 точку $M_2(x_2, y_2)$, где $x_2=x_1+\Delta x_1$, $y_2=y_1+\Delta y_1$ и $\frac{\Delta y_1}{\Delta x_1}=k_1=f(x_1, y_1)$, и т. д. Многократно повторяя эту процедуру, получаем некоторую ломаную (так называемая *ломаная Эйлера*, рис. 4), которая (при малых «шагах» Δx_0, Δx_1 и т. д.) дает достаточное представление о поведении *интегральной кривой* (т. е. решения) $y=y(x)$

уравнения (3) (ломаные Эйлера в определенном смысле «сходятся» к интегральной кривой, когда все их звенья стремятся к нулю). Соответствующие расчеты при современной вычислительной технике (компьютеры) оказываются достаточно простыми, а метод в целом — убедительным и надежным.

Другой метод решения дифференциальных уравнений основан на использовании степенных рядов (здесь, как и в случае ломаных Эйлера, речь идет, разумеется, вовсе не об одних лишь только уравнениях с разделяющимися переменными). Пусть мы имеем уравнение (3), в котором зависимость $f(x, y)$ от x и y достаточно проста (смысл этого последнего пока довольно туманного требования мы поясним чуть ниже), и хотим найти решение $y=y(x)$ этого уравнения, проходящее через фиксированную точку $M_0(x_0, y_0)$ плоскости. Предположим, что функция $y(x)$ в окрестности точки $x=x_0$, для которой $y(x_0)=a_0$, разложена в ряд Тейлора:

$$y(x)=a_0+a_1(x-x_0)+a_2(x-x_0)^2+ + a_3(x-x_0)^3+\ldots; \qquad (6.6.17)$$

тогда[1]

$$y'(x)=a_1+2a_2(x-x_0)+3a_3(x-x_0)^2+\ldots \qquad (6.6.18)$$

Подставляя выражения (17) и (18) в исходное уравнение (3), мы во многих случаях приходим к условию равенства двух степенных рядов: ряда (18) для y' и ряда, выражающего функцию $f(x, y)$, где y имеет вид (17) (вот тут-то и желательно, чтобы функция f была «достаточно проста», т. е. такова, что выражение $f(x,\ a_0+a_1(x-x_0)+a_2(x-x_0)^2+\ldots)$ можно представить как степенной ряд $b_0+b_1(x-x_0)+b_2(x-x_0)^2+\ldots$). Приравнивая коэффициенты этих двух рядов при одинаковых степенях x, мы часто можем из системы $a_1=b_0$, $2a_2=b_1$, $3a_3=b_2$ и т. д. численно определить коэффициенты $a_1, a_2, \ldots$ ряда (17), т. е. найти решение уравнения (3) (ведь коэффициент a_0 нам известен заранее)[2].

Для иллюстрации этого метода рассмотрим два простых примера, первый из которых может иметь лишь чисто иллюстративное значение, поскольку соответствующее уравнение легко решить и без рядов; второй же достаточно типичен.

1. Вернемся к уже рассмотренному выше уравнению (11)

$$y'=xy$$

и положим в нем

$$y=a_0+a_1x+a_2x^2+a_3x^3+\ldots; \qquad (6.6.19)$$

тогда

$$y'=a_1+2a_2x+3a_3x^2+4a_4x^3+\ldots \qquad (6.6.20)$$

Подставляя значения (19) и (20) в (11), получим

$$a_1+2a_2x+3a_3x^2+4a_4x^3+5a_5x^4+6a_6x^5+\ldots= =a_0x+\ {}_1x^2+a_2x^3+a_3x^4+a_4x^5+a_5x^6+\ldots \qquad (6.6.21)$$

Приравнивая теперь в (21) коэффициенты при одинаковых степенях x слева и справа, получаем:

$$a_1=0,\quad a_2=\frac{a_0}{2},\quad a_3=\frac{a_1}{3}=0,$$

$$a_4=\frac{a_2}{4}=\frac{a_0}{2\cdot 4},\quad a_5=\frac{a_3}{5}=0,$$

$$a_6=\frac{a_4}{6}=\frac{a_0}{2\cdot 4\cdot 6},\ \ldots,$$

где $a_0=y_0=y(0)$. В силу этих равенств разложение (19) в нашем случае имеет вид

$$y=a_0\left(1+\frac{x^2}{2}+\frac{1}{2}\frac{x^4}{4}+\frac{1}{2\cdot 3}\frac{x^6}{8}+\ldots\right),$$

откуда следует, что $y=a_0e^{x^2/2}$ (подставьте в правую часть формулы (2.2) вместо x величину $\frac{x^2}{2}$!).

2. Пусть

$$y'=xy+1. \qquad (6.6.22)$$

Уравнение (22) не является уравнением с разделяющимися переменными, и можно доказать, что его решение вообще нельзя записать в виде явной функциональной зависимости y от x. Несмотря на это, решение уравнения (22) нетрудно выразить в виде ряда.

Воспользуемся снова разложениями (19) и (20) и подставим их в (22). Мы получим

$$a_1+2a_2x+3a_3x^2+4a_4x^3+5a_5x^4+6a_6x^5+\ldots= =1+a_0x+a_1x^2+a_2x^2+a_3x^4+a_4x^5+\ldots,$$

откуда следует, что

$$a_1=1,\quad a_2=\frac{a_0}{2},\quad a_3=\frac{a_1}{3}=\frac{1}{1\cdot 3},$$

$$a_4=\frac{a_2}{4}=\frac{a_0}{2\cdot 4},\quad a_5=\frac{a_3}{5}=\frac{1}{1\cdot 3\cdot 5},$$

$$a_6=\frac{a_4}{6}=\frac{a_0}{2\cdot 4\cdot 6},\ \ldots$$

---

[1] Читателю придется принять на веру (не доказанную нами, но как будто достаточно естественную) возможность почленного дифференцирования степенных рядов, подобных ряду в правой части (17).

[2] Заметим, что разложение (17), как мы знаем, обычно справедливо лишь для некоторой окрестности точки $x=x_0$; поэтому зачастую приходится, дойдя до некоторого значения $x=x_1$, отбросить разложение (17) и пользоваться разложением y по степеням $x-x_1$ и т. д. Рассматриваемые ниже примеры выбраны так, чтобы избежать этих затруднений.

Таким путем мы получаем общее решение уравнения (22):

$$y=\left(\frac{x}{1}+\frac{x^3}{1\cdot 3}+\frac{x^5}{1\cdot 3\cdot 5}+\right.$$
$$\left.+\frac{x^7}{1\cdot 3\cdot 5\cdot 7}+\ldots\right)+a_0\left(1+\frac{x^2}{2}+\frac{x^4}{2\cdot 4}+\right.$$
$$\left.+\frac{x^6}{2\cdot 4\cdot 6}+\ldots\right), \qquad (6.6.23)$$

где значение a_0 определяется *начальным условием*: так, например, если мы знаем, что $y(0)=0$, то $a_0=0$ (ср. упр. 4).

Упражнения

6.6.1. Решите дифференциальные уравнения (с разделяющимися переменными): а) $y'=y^2$; б) $y'=k\frac{x}{y}$; в) $y'=k\frac{y}{x}$; г) $y'=x^2y^3$; д) $y'=\cos x\sin y$.

6.6.2. Докажите, что изображенные на рис. 2 прямые а) $y=0$; б) $y=\pm x$ также являются интегральными кривыми уравнений (11) и (15) соответственно.

6.6.3. Укажите «особое» решение уравнения (9), сходное с рассмотренными в упр. 2 «особыми» решениями уравнений (11) и (15).

6.6.4. Докажите, что проходящая через начало координат интегральная кривая (23) уравнения (22) может быть записана формулой

$$y=e^{x^2/2}\Phi(x), \text{ где } \Phi(x)=\int_0^x e^{-t^2/2}dt.$$

6.6.5. Решите методом разложения в ряд: а) уравнение (9); б) уравнение (13); в) уравнение из упр. 1а.

§ 7*. Дифференциальное уравнение вытекания воды

В качестве еще одного примера задачи на решение дифференциальных уравнений рассмотрим фигурировавший в § 3.6 вопрос о вытекании воды из сосуда. Мы будем считать, что рассматриваемый сосуд (ср. рис. 3.6.3) имеет внизу отверстие (или из него ведет тонкая трубка), в силу чего вода из сосуда может вытекать; однако наряду с этим теперь мы считаем, что в сосуд также может поступать вода из внешнего источника. Это задача очень проста и наглядна по своей постановке. Вместе с тем те же математические методы, которые позволяют описать рассматриваемый процесс вытекания воды, применяются и в более сложных и содержательных задачах.

Представим себе сосуд, в который втекает (или из которого вытекает) вода. Объем воды, находящейся в сосуде, обозначим через V. Этот объем со временем меняется, т. е. V есть функция времени t. Каков смысл величины $\frac{dV}{dt}$?

Ясно, что $dV\approx V(t+dt)-V(t)$ есть объем воды, поступившей в сосуд (при отрицательном dV — ушедшей из сосуда) за время dt. Поэтому $\frac{dV}{dt}=q(t)$ есть скорость изменения количества воды в сосуде. Величина $q(t)$ носит специальное название *потока* воды. Если $q>0$, то вода в сосуд поступает, если же $q<0$, то вода из сосуда вытекает, т. е. масса воды в сосуде уменьшается. В частном случае, когда поток q постоянен, не зависит от времени, $V(t+1)=V(t)+q$. Здесь q есть количество воды, поступившее в сосуд за единицу времени (оно может быть и отрицательным). В общем случае на протяжении единицы времени (за которую может быть принята не только 1 с, но и 1 ч или 1 год) величина q может меняться; поэтому-то и приходится определять q как производную. Для читателя, разобравшегося в соотношении между средней и мгновенной скоростью, между производной и приращением (см. гл. 2), дальнейшие разъяснения не нужны.

Мы знаем, что

$$\frac{dV}{dt}=q(t). \qquad (6.7.1)$$

Если зависимость потока воды от времени известна, т. е. известна функция $q(t)$, то дифференциальное уравнение (1) во всем подобно уравнению (6.4) и задача нахождения V математически не отличается от задачи определения пути по заданной скорости, которая, как мы знаем (см. гл. 3), решается с помощью вычисления (определенного) интеграла.

Для того чтобы наша задача имела определенное решение, нужно задать объем V_0 воды, который находился в сосуде в определенный начальный момент времени t_0. Условие $V=V_0$ при $t=t_0$ называется *начальным условием*, выделяющим одно определенное решение уравнения (1) (ср. с § 6).

Объем (количество) воды, которая втекла в сосуд (или ушла из него) за время от t_0 до t_1, есть $\int\limits_{t_0}^{t_1} q(t)\,dt$. Поэтому количество воды в сосуде в момент t_1 равно

$$V(t_1) = V_0 + \int\limits_{t_0}^{t_1} q(t)\,dt. \qquad (6.7.2)$$

Это выражение справедливо для любого момента времени t_1 и, следовательно, полностью определяет искомую зависимость V от t_1. При $t_1 = t_0$ интеграл в формуле (2) равен нулю и $V(t_0) = V_0$. Таким образом, решение (2) действительно удовлетворяет нашему начальному условию.

Формулой (2) можно пользоваться и при $t_1 < t_0$. Однако при $t_1 < t_0$ и при $t_1 > t_0$ эта формула имеет разный смысл. При $t_1 > t_0$ величина $V(t_1)$ есть количество воды, которое будет в сосуде в момент t_1, если объем воды в сосуде в момент t_0 был равен V_0 и поток воды задан функцией $q(t)$. При $t_1 < t_0$ величина $V(t_1)$ есть тот объем воды, который должен был находиться в сосуде в момент t_1 для того, чтобы в более позднее время, к моменту t_0, при потоке, заданном функцией $q(t)$, объем воды оказался равным V_0.

Вместо t_1 в (2) можно писать просто t, хотя такая запись и не вполне аккуратна, ибо t у нас — переменная интегрирования, но подобную неаккуратность мы допускали часто. В таком случае (2) принимает вид

$$V(t) = V_0 + \int\limits_{t_0}^{t} q(t)\,dt. \qquad (6.7.3)$$

Надо только помнить, что $q(t)$ здесь — это не значение q на верхнем пределе интегрирования, а (переменная) функция аргумента t, пробегающего все значения от t_0 до t; никаких недоразумений при этом не возникает.

Формулу (3), дающую решение задачи о вытекании воды, когда задан поток $q(t)$ и количество $V_0 = V(t_0)$ воды в начальный момент $t = t_0$, можно получить и при помощи несколько иных рассуждений. Из (1) в силу определения неопределенного интеграла следует

$$V(t) = \int q(t)\,dt.$$

Предположим, что первообразная функции $q(t)$ каким-либо образом найдена. Обозначим ее через $I(t)$. Тогда

$$\int q(t)\,dt = I(t) + C,$$

где C — постоянная интегрирования. Отсюда

$$V(t) = I(t) + C. \qquad (6.7.4)$$

Для определения постоянной интегрирования воспользуемся начальным условием, т. е. потребуем, чтобы при $t = t_0$ было $V = V_0$. Подставляя в (4) $t = t_0$, получим $V_0 = I(t_0) + C$, откуда $C = V_0 - I(t_0)$. Подставляя это значение C в (4), находим

$$V(t) = V_0 + I(t) - I(t_0),$$

что совпадает с формулой (3), поскольку

$$\int\limits_{t_0}^{t} q(t)\,dt = I(t)\Big|_{t_0}^{t} = I(t) - I(t_0).$$

Формулу (4) можно назвать *общим решением* дифференциального уравнения (1). Выбирая то или иное значение C, можно получить из формулы (4) различные *частные решения*, соответствующие различным начальным условиям.

Однако поток воды как функция времени известен отнюдь не всегда! Чаще известен физический закон, указывающий зависимость потока от напора воды, т. е. от высоты z уровня воды в сосуде (рис. 1). Так, например, при вытекании воды через тонкую длинную трубку можно считать, что

$$q = -kz, \qquad (6.7.5)$$

где коэффициент k — это некоторое положительное постоянное число, а знак минус означает, что вода в ы т е к а е т. При вытекании воды из отверстия в тонкой стенке имеет место совсем другой закон, установленный впервые учеником Галилея Э. Торричелли; здесь

$$q = -a\sqrt{z}. \qquad (6.7.5а)$$

Возможна также комбинация постоян-

ного поступления воды сверху (q_0) и вытекания ее по закону (5) или (5а):

$$q = q_0 - kz \text{ или } q = q_0 - a\sqrt{z}. \qquad (6.7.6)$$

В каждом из этих случаев, пока интересующая нас задача не решена, зависимость $z = z(t)$ уровня воды в сосуде от времени неизвестна, а значит, нам неизвестен и поток. Поэтому задачу определения из уравнения

$$\frac{dV}{dt} = q(z) \qquad (6.7.7)$$

функции V нельзя свести к предыдущей задаче.

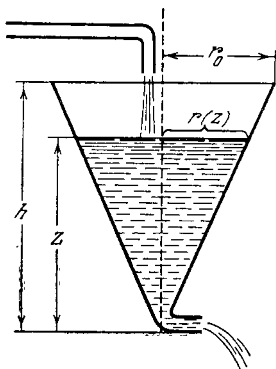


Рис. 6.7.1

Мы сформулировали здесь задачу в общем случае для произвольной зависимости $q = q(z)$ потока q от уровня z. В уравнение (7) входят две неизвестные величины: количество (объем) V воды и уровень воды z. Очевидно, эти величины не независимы: каждому уровню z соответствует вполне определенный объем V воды, так что V есть функция $V(z)$ переменной z.

Ясно, что вид функции $V(z)$ полностью определяется формой сосуда. Так, например, для цилиндрического сосуда радиуса r_0, если z есть расстояние от нижнего основания сосуда, то $V(z) = \pi r_0^2 z$. Для изображенного на рис. 1 конического сосуда по формуле для объема конуса находим $V(z) = {}^1/_3\, S(z)\, z$, где $S(z) = \pi[r(z)]^2$ — площадь сечения на высоте z (а $r(z)$ — радиус соответствующего сечения). Если r_0 — радиус верхнего основания конуса, а h — высота конуса, то из подобия треугольников легко получается, что $r(z) = \frac{r_0 z}{h}$; таким образом, здесь $V(z) = \frac{1}{3}\frac{\pi r_0^2}{h^2} z^3$.

Подставляя функцию $V = V(z)$ в уравнение (7), найдем

$$\frac{dV(z)}{dt} = \frac{dV(z)}{dz}\frac{dz}{dt} = q(z).$$

Производная $\frac{dV}{dz}$ от объема по высоте равна площади $S(z)$ сечения сосуда на высоте z (ср. формулу (3.6.20)): $\frac{dV(z)}{dz} = S(z)$. Окончательно получим

$$S(z)\frac{dz}{dt} = q(z). \qquad (6.7.8)$$

Ясно, что (8) есть дифференциальное уравнение с разделяющимися переменными. Рассмотрим теперь решения этого уравнения для разных видов зависимостей $S(z)$ и $q(z)$ [1].

Перепишем уравнение (8) в виде $\frac{dz}{dt} = \frac{q(z)}{S(z)}$ и обозначим $\frac{S(z)}{q(z)} = f(z)$. Тогда (8) примет вид

$$\frac{dz}{dt} = \frac{1}{f(z)}, \qquad (6.7.9)$$

а уравнения такого рода мы уже решали (ср. уравнение (6.4а)). Перепишем теперь (9) в виде

$$\frac{dt}{dz} = f(z). \qquad (6.7.10)$$

Запись (10) означает, что в ходе решения задачи мы временно будем рассматривать t как функцию z, т. е. будем искать не закон $z = z(t)$ изменения уровня воды, а обратную $z(t)$ функцию $t(z)$, — уже найдя $t(z)$, выразим z через t.

Умножим левую и правую части (10) на dz и проинтегрируем обе части получившегося равенства:

$$\int_{t_0}^{t} dt = \int_{z_0}^{z} f(z)\,dz. \qquad (6.7.11)$$

Отсюда имеем

$$t = t_0 + \int_{z_0}^{z} f(z)\,dz. \qquad (6.7.12)$$

Мы получили решение задачи: время t выражено как функция переменной z

---

[1] Разумеется, возможны и такие варианты задачи о вытекании воды (скажем, с резервуаром неопределенной формы, не позволяющим выписать явно функцию $S = S(z)$), когда соответствующее уравнение не допускает точного решения и его приходится решать приближенно (ср. с заключительной частью § 6, напечатанной мелким шрифтом).

(напомним, что t_0 — это просто число). Равенство (12) позволяет также при каждом значении t найти соответствующее ему z. Решение (12) удовлетворяет начальному условию $z=z_0$ при $t=t_0$ (в начальный момент $t=t_0$ задан уровень $z=z_0$ воды в сосуде).

Рассмотрим еще два конкретных примера зависимостей $f(z)$.

1. Вода вытекает из цилиндрического сосуда заданного радиуса r_0 по тонкой трубке. Здесь

$$S(z)=\text{const}=\pi r_0^2, \quad q=-kz,$$

так что уравнение (8) принимает вид

$$\pi r_0^2\frac{dz}{dt}=-kz, \qquad (6.7.13)$$

или

$$-\pi r_0^2\,\frac{dz}{z}=kdt.$$

Интегрируя левую часть в пределах от z_0 до z, а правую — в отвечающих этим значениям z пределах от t_0 до t, получим

$$-\pi r_0^2(\ln z-\ln z_0)=-\pi r_0^2\ln\frac{z}{z_0}=$$
$$=\pi r_0^2\ln\frac{z_0}{z}=k(t-t_0). \qquad (6.7.14)$$

Из (14) нетрудно выразить z через t. Действительно,

$$\ln z=\ln z_0-\frac{k}{\pi r_0^2}(t-t_0),$$

т. е.

$$z=z_0\exp\left[-\frac{k}{\pi r_0^2}(t-t_0)\right].$$

Рассмотрим два момента времени t и $t+\Delta t$ и найдем отношение $\frac{z(t+\Delta t)}{z(t)}$:

$$\frac{z(t+\Delta t)}{z(t)}=\exp\left[-\frac{k}{\pi r_0^2}(t+\Delta t-t_0-t+\right.$$
$$\left.+t_0)\right]=\exp\left(-\frac{k\Delta t}{\pi r_0^2}\right).$$

Мы видим, что это отношение зависит только от Δt и не зависит от t. Поэтому за равные промежутки времени уровень воды уменьшается в одном и том же отношении; он все время падает, экспоненциально убывая, но никогда не достигнет значения $z=0$, характеризующего полное вытекание воды [2].

Сравнительно мало меняет физическую картину и предположение о постоянной подаче воды в резервуар, когда $q(z)=q_0-kz$, и наше уравнение принимает вид

$$\pi r_0^2\frac{dz}{dt}=q_0-kz. \qquad (6.7.15)$$

Здесь удобно начать с нахождения *стационарного режима*, т. е. искать решение $z=z_{\text{ст}}=\text{const}$ рассматриваемого уравнения. Из предположения $z=z_{\text{ст}}$ и $\frac{dz}{dt}=0$ следует

$$q_0-kz_{\text{ст}}=0, \quad z_{\text{ст}}=\frac{q_0}{k}.$$

Ясно, что если процесс начинается со значения $z=z_{\text{ст}}$ уровня воды, то эта величина z не изменится: здесь подача воды в резервуар полностью компенсирует ее вытекание и уровень воды в сосуде остается постоянным. Пусть теперь начальное значение $z_0>z_{\text{ст}}$, так что (пропорциональное z) вытекание воды превышает ее поступление и уровень z уменьшается. Будем искать разность $z_1=z-z_{\text{ст}}$ уровня воды от стационарного уровня $z_{\text{ст}}$. Эта разность, очевидно, удовлетворяет уравнению, получаемому из (15) подстановкой значений $z=z_1+z_{\text{ст}}$ и $q_0=kz_{\text{ст}}$:

$$\pi r_0^2\frac{d(z_1+z_{\text{ст}})}{dt}=kz_{\text{ст}}-k(z_1+z_{\text{ст}}),$$

т. е. $\pi r_0^2\frac{dz_1}{dt}=-kz_1$.

Но это уравнение отличается от (13) лишь обозначениями, так что его решение имеет вид

$$z_1=z_1^{(0)}\exp\left[-\frac{k}{\pi r_0^2}(t-t_0)\right],$$

где $z_1^{(0)}=z_0-z_{\text{ст}}$. Окончательно получаем

$$z=z_{\text{ст}}+(z_0-z_{\text{ст}})\exp\left[-\frac{k}{\pi r_0^2}(t-t_0)\right];$$

здесь величина z стремится к стационарному уровню $z_{\text{ст}}$ воды, никогда его не достигая [3].

2. Вода вытекает из конического сосуда (с высотой h и

---

[2] Некоторая парадоксальность последнего заключения (вода из сосуда не вытечет никогда) связана с тем, что уравнение (13) относительно точно описывает рассматриваемый физический процесс лишь при $z\gg\rho$, где ρ — радиус трубки, через которую вытекает вода.

[3] Последнее заключение верно и при $z_0<z_{\text{ст}}$ (см. упр. 3), что свидетельствует об *устойчивости* установившегося состояния $z=z_{\text{ст}}$ воды в сосуде (ср. с § 9.3).

радиусом основания r_0; см. рис. 1) по тонкой трубке:

$$S(z)=\pi r_0^2\left(\frac{z}{h}\right)^2, \quad q=-kz.$$

В этом случае мы имеем уравнение

$$\frac{dz}{dt}=\frac{q(z)}{S(z)}=-\frac{kzh^2}{\pi r_0^2 z^2}=-\frac{kh^2}{\pi r_0^2}\frac{1}{z}, \qquad (6.7.16)$$

которое можно переписать так:

$$-\frac{\pi r_0^2}{kh^2}zdz=dt.$$

Интегрируя, получаем

$$-\frac{\pi r_0^2}{kh^2}\int\limits_{z_0}^{z}zdz=-\frac{\pi r_0^2}{kh^2}\left(\frac{z^2}{2}-\frac{z_0^2}{2}\right)=t-t_0.$$

И в этом примере не представляет труда выразить z как функцию t:

$$\frac{z^2}{2}=\frac{z_0^2}{2}-\frac{kh^2}{\pi r_0^2}(t-t_0),$$

или

$$z=\sqrt{z_0^2-\frac{2kh^2}{\pi r_0^2}(t-t_0)}. \qquad (6.7.17)$$

Формула (17) полностью решает задачу. Легко проверить, что

$$\frac{dz}{dt}=-\frac{kh^2}{\pi r_0^2}\left[z_0^2-\frac{2kh^2}{\pi r_0^2}(t-t_0)\right]^{-1/2}=-\frac{kh^2}{\pi r_0^2 z},$$

так что z действительно удовлетворяет уравнению. Очевидно также, что $z=z_0$ при $t=t_0$. Выражение (17) позволяет найти момент, когда сосуд полностью опорожнится:

$$z=0 \quad \text{при} \quad t=t_0+\frac{\pi r_0^2}{kh^2}\frac{z_0^2}{2}. \qquad (6.7.18)$$

Существенное качественное отличие первого примера от второго состоит в том, что во втором случае вся вода полностью выливается из сосуда за конечный интервал времени $T=\frac{\pi r_0^2}{kh^2}\frac{z_0^2}{2}$ (см. (18)), в то время как в первом случае мы имели лишь «асимптотическое» (продолжающееся неограниченно долго) приближение уровня воды к нулевой отметке $z=0$.

Упражнения

6.7.1. Рассмотрите решение уравнения (15), удовлетворяющее начальному условию $z_0 < z_{ст}$. Какой характер будет иметь в этом случае рассматриваемый процесс вытекания воды?

6.7.2. Рассмотрите случай вытекания воды по тонкой трубке из конического сосуда в условиях постоянного по интенсивности притока воды. [Указание. Дифференциальное уравнение рассматриваемого процесса имеет вид $\pi r_0^2\frac{z^2}{h^2}\frac{dz}{dt}=q_0-kz$.]

* * *

Если в главах 2 и 3 рассказывается об идейной стороне дифференциального и интегрального исчисления, то главы 4—6 посвящены аппаратной стороне (так сказать, «кухне») высшей математики — и каждому, кто хочет овладеть математическим анализом в такой степени, чтобы иметь возможность применять на практике соответствующие понятия и теоремы, необходимо эту аппаратную часть освоить. Конечно, главная задача, которая стоит перед естествоиспытателем или инженером, — это задача адекватного описания интересующего его процесса в математических терминах (чаще всего — в виде некоторого дифференциального уравнения); дальнейший чисто математический анализ модели (решение дифференциального уравнения) может быть поручен математику или передан программисту для машинной (использующей ЭВМ) обработки всего материала. (Но вот само построение математической модели явления перепоручить никому нельзя: это необходимо делать человеку, глубоко знакомому с разбираемым явлением!) Исходя из этого, инженер может решить, что в овладении математической техникой ему сегодня нет надобности, — и глубоко ошибется. Понимание сути высшей математики неразрывно связано с владением определенными техническими навыками: без минимальной техники не может быть и понимания принципиальной стороны дела. Так, для чтения специальной литературы на незнакомом языке, хотя бы и с помощью словаря (играющего в данном примере роль «программиста»), необходимо как владеть основными фактами грамматики («теории» языка), так и обладать некоторым минимальным запасом слов (начала «технических» знаний). Без представления о грамматическом строе языка мы будем беспомощны, не зная, какие именно слова (и в какой

форме) надо смотреть в словаре; но и без минимального запаса слов словарь не сможет нам помочь, ибо мы в растерянности остановимся перед совершенно темным для нас текстом. Однако уже небольшой запас слов и выражений плюс азы грамматики позволяют совершенно свободно пользоваться словарем, подобно тому как физику или инженеру зачастую (но не всегда!) может хватить знания начал высшей математики и владения простейшими техническими навыками.

Главы 4 и 5 книги дают общее (по необходимости — довольно беглое) представление о технике дифференцирования (нахождения производных) и интегрирования. Техника дифференцирования весьма проста, поэтому желательно отработать ее до той степени автоматизма, когда нахождение производных несложных функций осуществляется почти «само по себе». При этом вовсе не надо стремиться к решению большого числа сложных задач: ведь со сложными функциями физику или инженеру приходится встречаться не так уж часто. Гораздо важнее умение быстро решать простые задачи, поэтому читателю можно порекомендовать прорешать все собранные в настоящей книге упражнения на вычисление производных, но не увлекаться дифференцированием более сложных функций. Если решения задач из книги будет недостаточно для приобретения необходимых навыков, то можно обратиться к любому другому задачнику по дифференциальному исчислению или даже выписать самому ряд функций, так сказать, наугад и затем искать их производные. Пытаться запомнить сразу всю таблицу производных (см. Приложение I в конце книги) не следует: целесообразно на первой стадии работы переписать все формулы на карточку и пользоваться составленной таблицей в процессе решения задач. При этом формулы, к которым часто придется обращаться, запомнятся сами собой, и в конце концов табличка станет ненужной.

Гораздо сложнее техника интегрирования, изложенная в гл. 5. Нам кажется, что отработке этой техники не следует уделять слишком много внимания. Разумеется, полезно уметь без труда вычислять простые интегралы; однако, возможно, более целесообразно научиться свободно пользоваться таблицами интегралов. Но, конечно, умение интегрировать простейшие функции и приобретаемое в процессе работы с интегралами знание основных свойств интегралов (скажем, свободное пользование методом подстановки) необходимы каждому: ведь интегралы, с которыми Вы можете встретиться в своей работе, чаще всего не будут точно совпадать с теми, которые есть в таблицах, а лишь будут приводиться к табличным путем несложных преобразований, включающих и замену переменных.

Глава 6 начинается с темы о *рядах*. Техника замены сложных (например, экспериментально полученных) функций более простыми является одним из основных вкладов высшей математики в инженерную практику; представление о ней необходимо каждому физику и инженеру. Ключевое место здесь занимает учение о разложении функции в степенные ряды (см. § 6.1—6.4), хорошо дополняемое применяемым в других случаях разложением функций в так называемые «тригонометрические ряды», т. е. приближенным представлением их в виде простых комбинаций тригонометрических функций (см. § 10.9).

В настоящей книге мы ограничились лишь элементами соответствующей теории, переживающей ныне новый расцвет, связанный с целым рядом причин: бурным развитием тех разделов чистой (или теоретической) математики, которые широко используются в теории приближения функций (функциональный анализ, теория функций комплексного переменного и др.); запросами практики, постоянно встречающейся с необходимостью упрощения сложных функций; созданием ЭВМ, доставляющих новые возможности упрощенного представления функций и в то же время предъявляющих новые требования к вопросам задания функций в подходящем для использования в ЭВМ виде, и т. д.

Вторая тема, затронутая в главе 6 книги, связана с *дифференциальными уравнениями*, представляющими собой основной аппарат естествоиспытателя и инженера. В самом деле, математический анализ явлений природы обычно начинается с попыток представить те

или иные естественнонаучные законы в виде дифференциальных уравнений; эти уравнения связывают (переменные) величины, с помощью которых описывается интересующее нас явление (такое представление зачастую является не «абсолютным», а доставляет лишь приближенное описание реальной картины). В § 6.7 мы демонстрируем эту общую схему на примере изучения вытекания воды из сосуда. Заметим только, что как содержание § 6.7, так и разобранные в § 6.6 примеры дифференциальных уравнений имеют лишь иллюстративное значение: они свободно могут быть заменены любыми другими, и запоминать их не следует (хотя само понятие дифференциального уравнения первого порядка с разделяющимися переменными достаточно полезно для того, чтобы им следовало овладеть). В этом отношении уместно подчеркнуть различие между гл. 4 и последними параграфами гл. 6: если материал гл. 4 должен быть полностью усвоен каждым читателем книги, то § 6.6 и особенно § 6.7 достаточно бегло просмотреть, с тем чтобы понять принципиальную сторону вопроса.

Высшая математика была создана в XVII в. Ньютоном и Лейбницем, причем уже в их трудах она предстала как весьма содержательная дисциплина; однако развитие ее этим, конечно, не окончилось. Ньютон и Лейбниц отнюдь не ограничились основными определениями и первыми теоремами — их вклад был заметно бóльшим. Ньютон прекрасно понимал значение *дифференциальных уравнений* для анализа явлений природы: ведь его знаменитые «Математические начала натуральной философии» (1687), содержащие ньютонову механику (ньютонову систему мира), начинаются с дифференциального уравнения движения (см. (9.4.2)). Это уравнение принимается за аксиому, в то время как все последующие предложения механики являются теоремами, выводимыми из этой аксиомы (и из закона тяготения, следующего из экспериментальных фактов (законы Кеплера) и аксиомы (9.4.2)). В математических работах Ньютона сформулированы «основные задачи математического анализа»:

1) по данному соотношению между флюентами (исходными функциями) определить соотношение между флюксиями (производными);

2) по данному уравнению, содержащему флюксии, найти соотношение между флюентами.

Первая из этих задач есть, очевидно, задача дифференцирования известных комбинаций функций: так, в обозначениях Ньютона, если $z=uv$, то $\dot{z}=u\dot{v}+\dot{u}v$, здесь u, v и z — флюенты, а $\dot{u}$, $\dot{v}$, $\dot{z}$ — флюксии. Вторая же задача — это задача решения дифференциальных уравнений.

Много внимания уделял Ньютон и бесконечным *рядам* (открытие биномиального ряда произвело на него сильнейшее впечатление [4]): вся ньютонова теория флюксий и флюент (производных и интегралов) родилась, по-видимому, из его исследований по теории (бесконечных) рядов. И не случайно основополагающие теоремы (см. (1.18) и (1.19)) теории рядов связываются с именами непосредственных учеников и сотрудников Ньютона: Брука Тейлора (1685—1731), который был секретарем Лондонского Королевского общества (Британской академии наук) в те годы, когда Ньютон был его президентом, и профессора Эдинбургского университета Колина Маклорена (1698—1746), который был лично знаком с Ньютоном и являлся преданным его поклонником. Заметим кстати, что обе формулы (1.18) и (1.19) были впервые выписаны Тейлором; заслуга же Маклорена состояла в том, что он поставил вопрос об области применимости этих формул и частично ответил на него. Точная («конечная» — включающая конечное число слагаемых) формула (1.16) была установлена Ж. Л. Лагранжем, о котором мы еще скажем ниже.

Весьма серьезны и достижения Лейбница в теории рядов и в области дифференциальных уравнений, связанные с решением разнообразных геометрических задач. Напомним, что Лейбниц больше интересовался геометрией, чем механикой, и понятие производной ассоциировалось у него не со скоростью, а с касательной — недаром его основной мемуар по «новой математике» имел довольно сложное название: «Новый метод максимумов и минимумов, а также касательных, для которого не служат препятствием ни дробные, ни иррациональные величины, и особый для этого род исчисления».

Особо важное значение для дальнейшего прогресса дифференциального и интегрального исчисления имела первоклассная школа Лейбница, возглавляемая швейцарцами братьями Бернулли — Якобом (1654—1705) и Иоганном (1667—1748), родо-

---

[4] Относительно путей, какими, видимо, было сделано это выдающееся открытие, см., например: *Пойа Д.* Математическое открытие. М.: Наука, 1976, с. 119—121.

начальниками знаменитого «математического семейства» Бернулли, давшего в XVII—XVIII вв., восемь первоклассных математиков. (Самым значительным из последующих представителей этого семейства был работавший в Базеле и Петербурге сын Иоганна Даниил Бернулли (1700—1782) — один из основоположников гидродинамики и кинетической теории газов; с его именем мы в этой книге еще встретимся.)

Якоб Бернулли получил богословское образование, но тяга к математике возобладала: он самостоятельно ознакомился с математической литературой, в том числе с оказавшими на него глубочайшее впечатление работами Лейбница. Испытанное при чтении этих работ потрясение и привело его к отказу от обеспеченной карьеры пастора; ряд лет он занимал мало почетное и плохо оплачиваемое положение домашнего учителя, и лишь поддержка Лейбница помогла ему занять в 1683 г. должность профессора (сначала физики, а затем и математики) Базельского университета. Иоганна Бернулли отец хотел направить в область коммерции, — но и здесь желания отца пришли в противоречие с явно выраженными научными интересами Иоганна, основу которых заложил его старший брат, занимавшийся с ним математикой и физикой. Однако занятия математикой в Швейцарии тех лет давали мало возможностей для обеспеченного существования; поэтому И. Бернулли вынужден был получить медицинское образование и много лет зарабатывал на жизнь как медик (его докторская диссертация, посвященная движению мускулов, замечательна как, видимо, самый первый пример применения в физиологии методов высшей математики). В 1695 г. Иоганн по рекомендации знаменитого X. Гюйгенса получил профессуру по физике в Гронингене (Голландия); в 1705 г. он вернулся в Базель, где сначала смог получить в университете лишь место профессора греческого языка. Только после смерти Якоба Бернулли Иоганн занял в Базельском университете место профессора математики, которое и занимал до своих последних дней.

Братья Бернулли активно переписывались с Лейбницем, неоднократно восхищавшимся их успехами, — и именно в этой переписке математический анализ получил свою настоящую форму, основную символику и терминологию. Так, например, Лейбниц первоначально склонен был говорить о «дифференциальном исчислении» и «сумматорном исчислении» как о двух ветвях «новой математики», но по предложению Иоганна Бернулли остановился



ЯКОБ БЕРНУЛЛИ



ИОГАНН БЕРНУЛЛИ

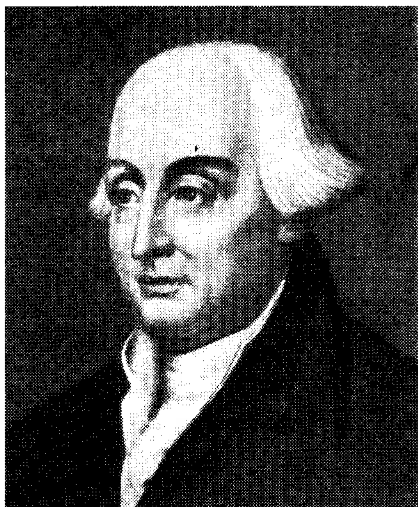
в конце концов на латинизированном термине «интегральное исчисление» (вместо «сумматорного»). Якоб и Иоганн Бернулли далеко продвинули новое исчисление, получив, в частности, важные результаты в области учения о дифференциальных уравнениях (см. § 6.6) и заложив основы так называемого *вариационного исчисления*, о котором мельком сказано в § 7.2 (см. текст, напечатанный на с. 189 мелким шрифтом [5]).

Первый печатный учебник дифференциального и интегрального исчисления «Анализ

[5] В самостоятельную научную дисциплину вариационное исчисление обратилось в XVIII веке благодаря трудам Л. Эйлера и Ж. Л. Лагранжа.



ЛЕОНАРД ЭЙЛЕР



ЖОЗЕФ ЛУИ ЛАГРАНЖ



ЖАН ЛЕРОН Д'АЛАМБЕР

бесконечно малых для исследования кривых линий» (обратите внимание на название, в котором слышны отголоски знаменитого мемуара Лейбница) вышел в свет в 1696 г.; автором его был ученик Лейбница и И. Бернулли французский дворянин маркиз Гийом Франсуа Антуан де Лопиталь (1661—1704). Этот замечательный учебник выдержал много переизданий и переведен на ряд языков (в том числе и на русский). С идейной стороны книга Лопиталя целиком следует лекциям Иоганна Бернулли, которые последний читал Лопиталю в Париже, и рукописному учебнику И. Бернулли «Лекции по исчислению дифференциалов», обнаруженному в библиотеке Базельского университета уже в нашем столетии и изданному лишь в 1922 г. Таким образом, в свое время учебник Бернулли оставался неизвестным, однако через посредство основанного на нем лекционного курса, который посещал всего один слушатель (маркиз Лопиталь), книга эта оказала огромное влияние на все дальнейшее развитие высшей математики (и, в частности, способствовала широкому распространению символики и терминологии Лейбница). В учебнике Лопиталя, в частности, был впервые опубликован сообщенный Иоганном Бернулли своему ученику метод вычисления пределов, ныне несправедливо называемый просто «правилом Лопиталя» (см. § 6.5).

Другим учеником Иоганна Бернулли был швейцарец Леонард Эйлер (1707—1783), которого познакомил с Бернулли отец Леонарда, пастор Пауль Эйлер, сам некогда изучавший математику под руководством Якоба Бернулли. Отец Эйлера хотел, чтобы и Леонард стал пастором; однако И. Бернулли сумел убедить Пауля Эйлера в выдающихся математических способностях сына. По рекомендации И. Бернулли Л. Эйлер в 1727 г. переехал в Петербург, где в недавно созданной Академии наук уже работали два сына его учителя (в том числе упоминавшийся выше Даниил Бернулли). Сначала предполагалось, что Эйлер займет имевшуюся в Петербургской академии наук вакансию профессора физиологии, и Эйлер глубоко изучил эту науку, готовясь по примеру своего наставника применить в ней математические методы. Но по приезде в Петербург оказалось, что свободно и место профессора математики. И далее Эйлер не возвращался к физиологии, зани-

маясь лишь математикой, физикой, механикой и астрономией (в частности, движениями Луны).

В Петербурге Эйлер и прожил вплоть до смерти с 25-летним перерывом, во время которого он по приглашению прусского короля Фридриха II перебрался из Петербурга в Берлин, где он руководил физико-математическим отделением Берлинской (Прусской) академии наук; впрочем, и в эти годы Эйлер не прерывал тесных связей с Петербургской академией. Эйлер был самым крупным и самым продуктивным математиком XVIII столетия; его работы касаются буквально всех областей математики и современной ему математической физики. С некоторыми из результатов Эйлера мы еще встретимся ниже (см. § 10.7 и гл. 14, 15). Принадлежащий Эйлеру многотомный учебник дифференциального и интегрального исчисления переведен и на русский язык.



ОГЮСТЕН КОШИ

Современником Эйлера был и крупный французский математик и механик Жан Лерон Д'Аламбер (1717—1783) — ученый, широчайшая научная культура которого ярче всего проявилась в создании знаменитой «Энциклопедии наук, искусств и ремесел», которую он выпускал вместе с Дени Дидро (1713—1784) и в которой ему принадлежали практически все статьи естественнонаучного содержания. Свое имя Д'Аламбер получил по названию церкви (Жан ле Рон — в переводе «Иоанн Круглый»), на ступенях которой был обнаружен мальчик-подкидыш, впоследствии, несмотря на все трудности, выросший в крупнейшего ученого. В § 10.7 мы расскажем об одной весьма поучительной и плодотворной научной дискуссии, в которой участвовали Л. Эйлер, Ж. Л. Д'Аламбер и Д. Бернулли. Заметим, что и Эйлера, и Д'Аламбера отличало стремление к получению конкретных результатов, выдающаяся математическая и физическая интуиция и отсутствие интереса к схоластическому обсуждению сущности математических понятий; с этим связан и известный призыв Д'Аламбера: «Работайте, работайте, — полное понимание придет потом».



БЕРНХАРД РИМАН

Выше мы уже упоминали, что с 1741 по 1766 г. Эйлер руководил физико-математическим отделением («классом») Берлинской академии наук. Когда он решил снова вернуться в Петербург, встал вопрос о его преемнике. Эйлер рекомендовал на свое место молодого математика, выходца из французской семьи, некогда перебравшейся в Италию, Жозефа Луи Лагранжа (1736—1813), которому тогда только исполнилось 30 лет; эта кандидатура была с энтузиазмом поддержана и Д'Аламбером, переписывавшимся с прусским королем Фридрихом II. Лагранж был в эти годы уже широко известным ученым: в 1726 г., в возрасте 20 лет, он начал преподавательскую работу в Туринской артиллерийской школе, а в следующем году стал одним из основателей и членом Туринской академии наук, в записках которой было опубликовано множество его статей. В 1787 г. после смерти Фридриха II Лагранж переехал в Париж, где сыграл выдающуюся роль в становлении парижской Политехнической школы — высшего учебного заведения нового типа, —

готовящей инженеров-исследователей [6]. К парижскому периоду жизни Лагранжа относится составление им двухтомной «Аналитической механики» (1788), сыгравшей огромную роль в дальнейшем прогрессе математики и физики. Лагранжу принадлежит также превосходный и во многом революционный учебник математического анализа («Теория аналитических функций», 1797), основанный на читавшихся им в Политехнической школе лекциях, а также основополагающие исследования по алгебре. По широте и разносторонности своих математических интересов и яркости достижений Лагранж из всех математиков XVIII в. уступал разве что Эйлеру.

Учеником и восторженным поклонником Лагранжа был профессор парижской Политехнической школы Жан Батист Фурье (1768—1830), с именем которого связаны выдающиеся достижения в области *дифференциальных уравнений в частных производных* (называемых еще *уравнениями математической физики*, ср. § 10.7), в частности в теории распространения тепла, и в теории тригонометрических рядов (см. § 10.8).

Профессором Политехнической школы был и знаменитый Огюстен Луи Коши (1789—1857), создатель современной теории пределов (основные идеи которой, впрочем, были заимствованы им у Д'Аламбера), способствовавший уточнению всех понятий высшей математики. Коши также по праву считается создателем *теории функций комплексной переменной* (см. гл. 14, 15 и 17); впрочем, эту честь он делит с одним из крупнейших ученых XIX в. — немецким математиком Георгом Фридрихом Бернхардом Риманом (1826—1866), имевшим выдающиеся результаты буквально во всех областях математики и математической физики.

Если Коши представлял в науке парижскую Политехническую школу, то Риман был связан с другим учебным и научным центром, сыгравшим большую роль в развитии в XIX и XX вв. математики и физики, — с Гёттингенским университетом в Германии, где он слушал лекции знаменитого математика и физика, астронома и геодезиста Карла Фридриха Гаусса (1777—1855), обычно признаваемого первым математиком XIX в. В Гёттингенском университете Риманом был впервые прочитан курс теории функций комплексной переменной, а также курс теории дифференциальных уравнений в частных производных; опубликованные уже после смерти Римана лекции по этому курсу явились первым учебником по уравнениям математической физики.

К совсем другому поколению гёттингенских ученых относится Давид Гильберт (1862—1943), считающийся создателем *функционального анализа*, изучающего «функции от функции», или функционалы (ср. с. 189); к этой области математики сегодня относят и учение об «обобщенных функциях», подобных дельта-функции Дирака (см. гл. 16 и 17).

Интенсивная разработка идей математического анализа продолжается и в наши дни; при этом создание ЭВМ придало новые мощные импульсы научной работе в этом направлении.

[6] Пример прославленной Политехнической школы Гаспара Монжа (1746—1818) и Ж. Л. Лагранжа, безусловно, учитывался и при создании Московского физико-технического института, а также американских Массачусетского и Калифорнийского технологических институтов.

Глава 7

ИССЛЕДОВАНИЕ ФУНКЦИЙ. НЕСКОЛЬКО ЗАДАЧ ИЗ ГЕОМЕТРИИ

В гл. 1—6 мы ввели основные понятия высшей математики — понятия *производной* и *интеграла* — и наметили начала техники работы с ними. Настоящая глава, завершающая первую часть книги, частично продолжает и развивает материал предыдущих глав; так, например, здесь мы подробно останавливаемся на методике отыскания максимальных и минимальных значений функций, о чем выше говорилось лишь достаточно бегло (см. § 2.6). В этой главе мы расскажем также об основных геометрических приложениях дифференциального и интегрального исчисления. Некоторые из тем этой главы являются факультативными (необязательными), они рассчитаны только на любителей (см., например, § 3, 5, 6).

Поскольку эта глава имеет характер дополнений к главам 1—6, то она неизбежно получилась достаточно пестрой; отдельным ее «мотивам» посвящены § 1 и 2; 4, 8—10; 11; 3 и 7; 5; 6.

§ 1. Гладкие максимумы и минимумы

Задача о нахождении такого значения x_0, при котором данная функция $y=f(x)$ достигает максимума или минимума, неразрешима в общем виде средствами элементарной алгебры.

В гл. 2 мы установили, что *в точках, где функция достигает максимума или минимума, ее производная равна нулю* (теорема Ферма). Там же было показано, как, пользуясь производной y', выяснить, является ли данная точка x_0, такая, что $f'(x_0)=0$, для рассматриваемой функции точкой максимума, минимума или точкой перегиба, в которой не достигается ни максимум, ни минимум. Для этого приходилось вычислять значения y' при значениях x, близких к x_0 и расположенных справа и слева от x_0; так, скажем, условия $y'(x)>0$ при $x<x_0$; $y'=0$ при $x=x_0$; $y'(x)<0$ при $x>x_0$ означают, что в точке x_0 функция $y=y(x)$ достигает максимума. В гл. 2 указывался также и еще один способ исследования того же вопроса, при котором к рассмотрению привлекалась вторая производная y'' (однако требовалось знать лишь ее значение в точке $x=x_0$). А именно, мы видели, что, если, например, $f'(x_0)=0$, $f''(x_0)<0$, то в точке $x=x_0$ функция $f(x)$ имеет максимум.

Последнее можно усмотреть и из соображений, связанных с понятием выпуклости графика функции; этого понятия мы касались в § 2.7 и еще специально займемся им в § 3. Действительно, условие $f'(x_0)=0$ означает, что касательная к графику функции в точке $x=x_0$ горизонтальна. Из неравенства $f''(x_0)<0$ следует, что точка $x=x_0$ является *точкой выпуклости*, т. е. график вблизи $x=x_0$ расположен под касательной (ср. рис. 2.7.1). Но эти два факта и означают, что функция $f(x)$ в точке $x=x_0$ имеет максимум. Аналогичные же рассуждения показывают, что если $f'(x_1)=0$, $f''(x_1)>0$, то в точке $x=x_1$ функция $f(x)$ имеет минимум.

Эти выводы подтверждаются также рассмотрением ряда Тейлора

$$f(x)=f(x_0)+f'(x_0)(x-x_0)+\frac{1}{2!}f''(x_0)(x-x_0)^2+\ldots \qquad (7.1.1)$$

функции $f(x)$ в точке $x=x_0$. Если $f'(x_0)\neq 0$, то при x, близких к x_0, величинами $(x-x_0)^2$, $(x-x_0)^3$ и т. д. можно пренебречь по сравнению с $(x-x_0)$. Таким образом, получаем приближенные равенства:

$$f(x)\approx f(x_0)+f'(x_0)(x-x_0), \quad \text{или} \quad f(x)-f(x_0)\approx f'(x_0)(x-x_0),$$

из которых видно, что если, скажем, $f'(x_0)>0$, то $f(x)-f(x_0)>0$ (т. е. $f(x)>f(x_0)$) при $x>x_0$ и $f(x)-f(x_0)<0$ (т. е. $f(x)<f(x_0)$) при $x<x_0$. Значит, здесь при $x=x_0$ мы не имеем ни максимума, ни минимума. (Аналогично нет ни максимума, ни минимума при $f'(x_0)<0$, только в первом случае функция в точке x_0 растет, а во втором — убывает.)

Если же $f'(x_0)=0$, то член с $(x-x_0)$ в (1) выпадает и пренебречь членом с $(x-x_0)^2$ уже нельзя. Пренебрегая членами с $(x-x_0)^3$, $(x-x_0)^4$ и т. д., малыми по сравнению с $(x-x_0)^2$, получаем из (1)

$$f(x) \approx f(x_0)+\frac{1}{2!}f''(x_0)(x-x_0)^2 \qquad (7.1.2)$$

(приближенное равенство (2) выполняется тем точнее, чем меньше $|x-x_0|$). Отсюда следует, что если $f''(x_0)>0$, то $f(x)>f(x_0)$ независимо от того, будет $x<x_0$ или $x>x_0$, т. е. $f(x_0)$ меньше всех соседних значений $f(x)$ и, значит, $f(x_0)$ — минимальное значение функции. Если же $f''(x_0)<0$, то $f(x)<f(x_0)$ и $f(x_0)$ — максимальное значение функции.

Может случиться, что $f'(x_0)=0$ и одновременно $f''(x_0)=0$. В этом случае нам придется учесть следующие члены ряда Тейлора (1). Если $f'''(x_0)\neq 0$, то членом с $(x-x_0)^3$ уже пренебречь нельзя; однако в этом случае можно свободно отбросить члены с множителями $(x-x_0)^4$, $(x-x_0)^5$ и т. д., малыми по сравнению с $(x-x_0)^3$. Поэтому здесь

$$f(x) \approx f(x_0)+\frac{1}{6}f'''(x_0)(x-x_0)^3, \qquad (7.1.2а)$$

и мы не имеем при $x=x_0$ ни максимума, ни минимума, подобно тому как не имеет ни максимума, ни минимума при $x=0$ функция $y=x^3$ (см. рис. 1.5.1; для этой функции $y'(0)=y''(0)=0$). Если же, скажем, $f'(x_0)=f''(x_0)=f'''(x_0)=0$, но $f^{IV}(x_0)\neq 0$, то в окрестности точки x_0

$$f(x) \approx f(x_0)+\frac{1}{24}f^{IV}(x_0)(x-x_0)^4. \qquad (7.1.2б)$$

Здесь знак выражения $f(x)-f(x_0)$ один и тот же при $x<x_0$ и при $x>x_0$: он определяется знаком $f^{IV}(x_0)$. Если $f^{IV}(x_0)>0$, то мы имеем минимум функции (как в случае функции $y=x^4$; см. рис. 1.5.1), а если $f^{IV}(x_0)<0$, — максимум.

Внимательный читатель уже, вероятно, догадался, что если при $x=x_0$ первая не равная нулю производная есть производная нечетного порядка (первая, третья, пятая и т. д.), то в этой точке нет ни максимума, ни минимума; если же это производная четного порядка (вторая, четвертая и т. д.), то мы имеем либо максимум, либо минимум, в зависимости от знака производной (минимум, если эта производная положительна; максимум, если она отрицательна). Здесь, как и всюду в этом параграфе, мы ограничиваемся рассмотрением максимумов и минимумов функций, достигаемых в точках, в которых график функции представляет собой *гладкую кривую*, т. е. в точках, где функция *дифференцируема* — имеет производную (или даже несколько последовательных производных $y'(x_0)$, $y''(x_0)$ и т. д.). Такие максимумы и минимумы иногда называют *гладкими*; иные случаи, связанные с так называемыми негладкими максимумами и минимумами, будут рассмотрены в § 2.

Перейдем к примерам.

Пример 1. Пусть

а) $y=A+B(x-a)^3$;

б) $y=A+B(x-a)^4$, (7.1.3)

где $B\neq 0$; найти максимумы и минимумы функции y.

Последовательные производные функций (3) таковы:

а) $y'=3B(x-a)^2$, $y''=6B(x-a)$

(и $y'(a)=y''(a)=0$), $y'''=6B\neq 0$;

б) $y'=4B(x-a)^3$, $y''=12B(x-a)^2$,

$y'''=24B(x-a)$ (и $y'(a)=y''(a)=y'''(a)=0$), $y^{IV}=24B\neq 0$.

В случае а) первой отличной от нуля оказалась производная нечетного (третьего) порядка; поэтому здесь в точке $x=a$ нет ни максимума, ни минимума (и функция y их вовсе не имеет, ибо $y'(x)=0$ лишь при $x=a$), а есть точка перегиба. В случае б) отлична от нуля в точке $x=a$ производная четного (четвертого) порядка и функция y имеет минимум при $B>0$ ($y^{IV}(a)>0$) и максимум при $B<0$ ($y^{IV}<0$), что, впрочем, ясно и из формулы для функции y (почему?).

Пример 2. Из квадратного листа жести, сторона которого равна $2a$, требуется сделать открытый сверху ящик возможно большего объема, вырезая равные квадраты по углам, удаляя их и затем загибая жесть, чтобы образо-

вать бока ящика (рис. 1). Какова должна быть длина стороны у вырезаемых квадратов?

Пусть стороны вырезаемых квадратов равны x; объем ящика (он зависит от того, какой квадрат мы вырезаем) обозначим через $V(x)$. Ясно, что

$$V(x)=(2a-2x)^2x=4(a-x)^2x, \quad (7.1.4)$$

— ведь основанием (прямоугольного) ящика является квадрат со стороной $2a-2x$, а его высота равна x.

Найдем теперь производную функции (4):

$$V'(x)=-8(a-x)x+4(a-x)^2.$$

Решим уравнение $V'(x)=0$:

$$-8(a-x)x+4(a-x)^2=0,$$

или $(a-x)(a-3x)=0$,

откуда $x_1=a, \quad x_2=\frac{a}{3}$.

Значение $x_1=a$ нас, разумеется, не интересует: ведь при таком способе разрезания листа никакого ящика не получится. Остается значение $x=\frac{a}{3}$. При этом

$$V\left(\frac{a}{3}\right)=4\frac{4a^2}{9}\frac{a}{3}=\frac{16a^3}{27}\approx 0{,}593a^3.$$

Здесь $V'\left(\frac{a}{3}\right)=0$, а поскольку

$$V''(x)=8x-8(a-x)-8(a-x)=$$
$$=24x-16a,$$

то $V''\left(\frac{a}{3}\right)=-8a<0$. Следовательно, функция $V(x)$ при $x=\frac{a}{3}$ имеет максимум.

Итак, наибольшее значение объема ящика достигается при $x=\frac{a}{3}$, т. е. если вырезать квадратики, стороны которых составляют $\frac{1}{6}$ стороны исходного квадрата.

Подсчитаем $V(x)$ при нескольких x, близких к $\frac{a}{3}$:

x	$0{,}25a$	$0{,}30a$	$0{,}33a$	$0{,}40a$	$0{,}45a$
$V(x)$	$0{,}562a^3$	$0{,}588a^3$	$0{,}592a^3$	$0{,}576a^3$	$0{,}540a^3$

Видно, что вблизи $x=\frac{a}{3}$, т. е. вблизи значения x, которому соответствует максимум функции, малые изменения x вызывают весьма малые изменения V: функция вблизи максимума изменяется очень медленно.

Это же следует и из формулы Тейлора (1). Если $f'(x)=0$ (максимум или минимум функции), то ряд (1) принимает вид

$$f(x)=f(x_0)+\frac{1}{2}f''(x_0)(x-x_0)^2+$$
$$+\frac{1}{6}f'''(x_0)(x-x_0)^3+\ldots \quad (7.1.5)$$

Таким образом, он не содержит члена с $(x-x_0)$: наименьшая степень разности $x-x_0$ в правой части (5) есть $(x-x_0)^2$,

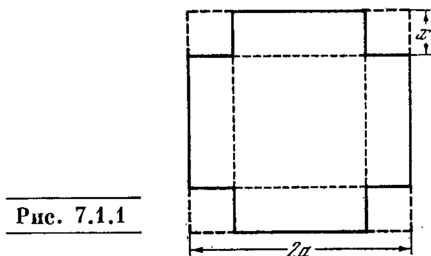


Рис. 7.1.1

и при x, близком к x_0, она будет весьма мала. В нашем примере изменение x на 9% (от $0{,}33a$ до $0{,}30a$) вызывает изменение V меньше чем на 1%, а изменение x на 24% (от $0{,}33a$ до $0{,}25a$) вызывает изменение V на 5%. Поэтому если нас интересует максимальное значение функции, а при нахождении x_0 из уравнения $V'(x)=0$ мы допустили небольшую ошибку (скажем, решили это уравнение приближенно), то это мало повлияет на величину найденного максимального значения функции.

Заметим еще, что в этой задаче то обстоятельство, что точка $x=\frac{a}{3}$ есть именно точка **максимума** нашей функции, можно усмотреть и просто из соображений здравого смысла, без обращения ко второй производной. Функция, выражающая объем ящика, очевидно, положительна при всех x, заключенных между 0 и a; при «крайних» же значениях $x=0$ и $x=a$ она обращается в нуль (в первом случае обращается в нуль высота ящика, а во втором — площадь его основания). Отсюда следует, что при изменении x

от 0 до a величина $V(x)$ сначала увеличивается, а затем начинает уменьшаться. Ясно, что при этом она где-то должна достигнуть максимума, а так

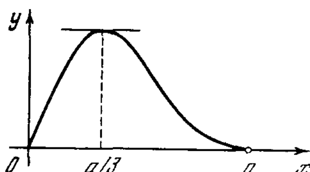


Рис. 7.1.2

как «подозрительной» в этом отношении является только точка $x=\frac{a}{3}$ (ибо лишь в этой точке производная $V'(x)$ равняется нулю), то она и является именно

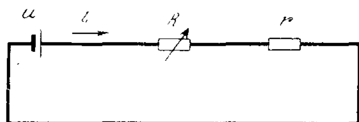


Рис. 7.1.3

точкой максимума. Для того чтобы сделать эти рассуждения более ясными, мы изобразили на рис. 2 график функции $V=V(x)$ (в точке $x=a$ график касается оси абсцисс, ибо $V'(a)=0$).

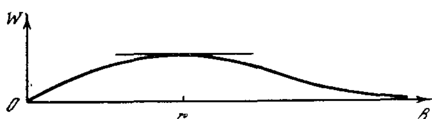
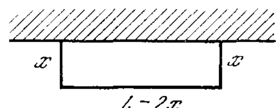


Рис. 7.1.4

Пример 3. Пусть мы имеем электрическую сеть, включающую тот или иной источник тока, создающий определенную ЭДС u (ср. с гл. 13), а также сопротивление r (возможно,

Рис. 7.1.5



внутреннее сопротивление источника напряжения); наряду с этим в цепь включено еще одно (переменное) сопротивление R (рис. 3). Требуется подобрать сопротивление R так, чтобы выделяемая на нем мощность W была максимальной.

Поскольку сопротивление цепи равно $r+R$, то сила i текущего по нашей цепи тока в силу закона Ома равна $i=\frac{u}{r+R}$. Мощность $W=iu_R$, где u_R — падение напряжения на сопротивлении R, по закону Ома равное iR. Подставляя это значение u_R в выражение для мощности W, находим

$$W=i^2R=\frac{u^2R}{(R+r)^2}. \qquad (7.1.6)$$

Нам надо найти максимум выражения (6) при переменном R. Продифференцируем это выражение по R:

$$\frac{dW}{dR}=\frac{u^2}{(r+R)^2}-\frac{2u^2R}{(r+R)^3}=$$
$$=\frac{u^2(r-R)}{(r+R)^3}. \qquad (7.1.7)$$

Из (7) следует, что $\frac{dW}{dR}=0$ при $R=r$.

Нетрудно понять, что значение $R=r$ будет отвечать именно максимуму функции $W(R)$; это следует хотя бы из того, что при $R<r$ производная (7) этой функции положительна, а при $R>r$ — отрицательна. Впрочем, максимальность энергии W при $R=r$ вытекает и из соображений здравого смысла: ведь при малом R и энергия W мала из-за малости сопротивления (в этом случае почти вся энергия «сидит» на сопротивлении r, а на сопротивление R почти ничего не приходится; если $R=0$, то и $W=0$); если же R очень велико, то в силу закона Ома будет очень малой сила тока i, а значит, и энергия W; в промежутке между малыми и большими R энергия W где-то должна достигнуть максимума, а может она достигнуть его лишь при $R=r$ (ср. с графиком функции $W=W(R)$ на рис. 4).

Пример 4. Из имеющихся досок можно сделать забор длиной l. Как этим забором огородить прямоугольный двор наибольшей площади, используя в качестве одной стороны стену прилегающего здания (рис. 5)?

Пусть две стороны забора имеют длину x; тогда третья сторона имеет длину $l-2x$. Площадь двора

$$S(x)=(l-2x)x=-2x^2+lx, \qquad (7.1.8)$$

откуда

$$S'(x)=-4x+l.$$

Ясно, что $S'(x)=0$ лишь при $x=\frac{l}{4}$; здесь $S''(x)=-4<0$, так что при $x=\frac{l}{4}$ функция $S(x)$ имеет максимум. (Впрочем, и здесь наличие максимума вытекает из соображений здравого смысла; см. конец разбора примера 2.)

Этот результат легко получить и не используя дифференциальное исчисление.

Действительно, нам нужно отыскать максимум (или минимум) квадратичной функции (8) переменной x. Но для любого многочлена второй степени

$$y=ax^2+bx+c \tag{7.1.9}$$

имеем

$$y=a\left(x^2+\frac{b}{a}x+\frac{c}{a}\right)=a\left[x^2+2\frac{b}{2a}x+\right.$$
$$\left.+\frac{b^2}{4a^2}-\frac{b^2}{4a^2}+\frac{c}{a}\right]=a\left[\left(x+\frac{b}{2a}\right)^2+\right.$$
$$\left.+\frac{4ac-b^2}{4a^2}\right]=a\left(x+\frac{b}{2a}\right)^2+\frac{4ac-b^2}{4a}. \tag{7.1.9a}$$

Так как $\left(x+\frac{b}{2a}\right)^2\geqslant 0$ при всех x $\left(\text{причем равенство нулю имеет место лишь при } x=-\frac{b}{2a}\right)$, то y имеет максимум, если $a<0$, и этот максимум достигается при $x=-\frac{b}{2a}$; y имеет минимум, если $a>0$, и этот минимум достигается при $x=-\frac{b}{2a}$. В частности, функция (8), где $a=-2$, $b=l$, $c=0$, достигает максимума при $x=\frac{l}{4}$.

Рассмотренная нами задача является одним из вариантов так называемой *задачи Дидоны*. По преданию, мифическая основательница города Карфагена в Северной Африке финикийская царица Дидона в ответ на обращенную к вождю прибрежного племени просьбу о выделении ей территории для постройки города получила издевательское согласие уступить участок земли «в пределах бычьей шкуры». Однако хитрая Дидона не просто покрыла шкурой крошечную часть побережья, как рассчитывали аборигены, а, разрезав шкуру на тонкие ремни, отгородила этими ремнями довольно большой участок, который удалось сделать еще большим, воспользовавшись берегом моря. Если считать берег прямолинейным и требовать, чтобы отгороженный участок был прямоугольным, то мы придем к решенной нами задаче. Задача оказывается гораздо более сложной, если предположить, что граница участка, т. е. лента из бычьей шкуры, может иметь произвольную форму. В этом случае площадь, максимум которой требуется найти, зависит не от одной переменной x, а от произвольной функции, задающей форму линии, отгораживающей требуемый участок. Однако и на этот случай удается перенести методы высшей математики,

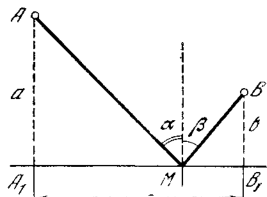
Рис. 7.1.6



в частности введя и здесь аналог понятия дифференциала и доказав, что точкам максимума и минимума рассматриваемых «функций от функций» (их математики называют *функционалами*) отвечает обращение в нуль этого «обобщенного дифференциала». С помощью таких рассмотрений (которых мы, разумеется, никак не можем коснуться в этой элементарной книге) удается доказать, что решением общей задачи Дидоны является полуокружность (рис. 6).

Пример 5. Турист должен пройти из палатки A к костру B, предварительно набрав воду в реке (прямая A_1B_1). Как проделать требуемый путь, пройдя наименьшее расстояние (рис. 7)?

Рис. 7.1.7



Пусть $AA_1=a$, $BB_1=b$, $A_1B_1=c$ ($AA_1 \parallel BB_1 \perp A_1B_1$; значения чисел a, b, c нам даны). Путь пешехода изображается ломаной AMB. Нужно узнать, при каком положении точки M на прямой A_1B_1 этот путь будет наименьшим. Положение точки M задается расстоянием M от точки A_1, т. е. от перпендикуляра AA_1, опущенного из A на прямую, изображающую реку; это расстояние A_1M мы обозначим через x. Тогда

$$AM=\sqrt{a^2+x^2},\quad MB=\sqrt{b^2+(c-x)^2}$$

и пройденный пешеходом путь $S(x)$ будет равен

$$S(x)=AM+MB=\sqrt{a^2+x^2}+$$
$$+\sqrt{b^2+(c-x)^2}. \tag{7.1.10}$$

Отсюда находим

$$S'(x) = \frac{x}{\sqrt{a^2+x^2}} - \frac{c-x}{\sqrt{b^2+(c-x)^2}}.$$

Приравняем $S'(x)$ нулю:

$$\frac{x}{\sqrt{a^2+x^2}} = \frac{c-x}{\sqrt{b^2+(c-x)^2}}. \quad (7.1.11)$$

Уравнение (11) нетрудно решить. Возведя обе части в квадрат, получим

$$\frac{x^2}{a^2+x^2} = \frac{(c-x)^2}{b^2+(c-x)^2},$$

или $x^2b^2+x^2(c-x^2)=a^2(c-x)^2+$ $+x^2(c-x)^2$,

т. е. $x^2b^2=a^2(c-x)^2$, $\frac{x^2}{(c-x)^2}=\frac{a^2}{b^2}$,

откуда

$$\frac{x}{c-x} = \pm\frac{a}{b},$$

или $x_1=\frac{ac}{a+b}$, $x_2=\frac{ac}{a-b}$.

Подставляя эти значения x_1 и x_2 в исходное уравнение (11), видим, что второй корень не удовлетворяет уравнению: этот посторонний корень получился из-за возведения обеих частей (11) в квадрат [1]. Итак,

$$x = \frac{ac}{a+b}.$$

Однако можно, не решая уравнения (11), выяснить его геометрический смысл, позволяющий получить ответ. Условие (11) можно переписать так:

$$\frac{A_1M}{AM} = \frac{MB_1}{MB}. \quad (7.1.11а)$$

Но $\frac{A_1M}{AM} = \cos A_1\widehat{M}A = \sin\alpha$; аналогично $\frac{MB_1}{MB} = \cos B_1\widehat{M}B = \sin\beta$. Соотношение (11а) означает, что $\sin\alpha = \sin\beta$, или, поскольку оба угла α и β острые, что

$$\alpha = \beta. \quad (7.1.12)$$

Таким образом, пешеход должен двигаться так, как движется отражающийся от зеркала луч света (*угол падения равен углу отражения*).

[1] Посторонний корень x_2 уравнения (11) можно сразу отбросить и потому, что, очевидно, $\frac{x}{c-x}>0$ и, значит, не равно $-\frac{a}{b}$.

Для полного решения задачи остается показать, что при таком положении точки M путь действительно будет минимальным (а не максимальным). Это можно сделать, вычислив вторую производную от (10).

Можно, однако, использовать и другие соображения. Из выражения (10) для $S(x)$ видим, что при любом x величина $S(x)$ положительна. При этом $S(x)$ неограниченно возрастает вместе с ростом абсолютной величины x независимо от того, будет $x>0$ или $x<0$. А так как $S'(x)$ обращается в нуль лишь при одном значении x, то ясно, что при этом значении x функция $S(x)$ имеет минимум. Вообще, если

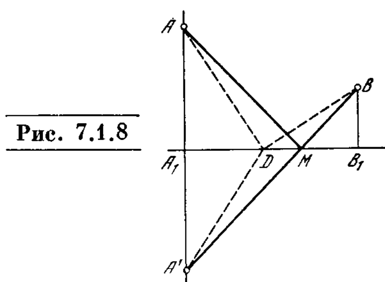


Рис. 7.1.8

в интересующем нас промежутке первая производная имеет лишь один корень, то наглядные соображения часто позволяют избежать формального исследования при помощи второй производной (ср. со сказанным по поводу примера 2).

Нашу задачу можно решить и чисто геометрически, не прибегая к методам высшей математики. На продолжении отрезка AA_1 (рис. 8) отложим $A_1A'=AA_1$ и точку A' соединим с B. Тогда $AM=A'M$, так как $\Delta AA_1M=\Delta A_1A'M$. Поэтому $A'B=A'M+MB=AM+MB$. Для любой другой точки D на отрезке A_1B_1 будет $AD+DB=A'D+DB$ и $A'D+DB>A'B$, так как ломаная длиннее отрезка прямой. Искомая точка M есть, следовательно, точка пересечения прямых $A'B$ и A_1B_1, откуда и вытекает равенство (12) (почему?).

Примеры 4 и 5 показывают, что задачи на нахождение максимумов и минимумов иногда удается решить и средствами элементарной математики, не прибегая к дифференциальному исчислению (см. решения соответствующих задач, напечатанные мелким шрифтом). Однако, во-первых, далеко не все задачи можно осилить, не используя высшую математику. Во-вторых, решение элементар-

ными средствами требует смекалки и остроумия; высшая же математика дает стандартный способ решения таких задач, не требующий никакой изобретательности.

Это конечно, вовсе не означает, что в высшей математике можно обойтись без сообразительности! Но она теперь пригодится для других, более трудных случаев.

Пример 6. Вот еще один, несколько более сложный, чем предыдущие, пример исследования функции на максимум и минимум. Хорошо известно, что связь между объемом v, давлением p и (абсолютной — по Кельвину) температурой T идеального газа задается законами Бойля — Мариотта и Гей-Люссака, которые объединяются в одном *уравнении Клапейрона*:

$$pv = RT, \tag{7.1.13}$$

где R — некоторая постоянная (*универсальная постоянная* идеального газа; прилагательное «универсальная» означает, что при отнесении закона (13) к одной грамм-молекуле газа величина R будет одна и та же для всех газов; так, если v измеряется в м^3, а p — в Па, то $R \approx 8{,}25$ Н·м/град). Однако формула (13) для реальных газов выполняется недостаточно точно, в связи с чем широко используется следующее уточнение уравнения (13):

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT, \tag{7.1.14}$$

называемое *уравнением ван дер Ваальса* [2]; здесь a и b — некоторые (экспериментально определяемые) положительные постоянные (при указанных выше единицах измерения имеющие размерность соответственно Н·м^4 и м^3), разные для различных газов. При постоянной температуре T формулы (13) и (14) описывают некоторые кривые в плоскости, отнесенной к координатам v и p; эти кривые называют *изотермами* (линиями одинаковых температур). Каждому значению T отвечает своя изотерма.

Ясно, что задаваемые формулой (13) «изотермы Клапейрона» представляют собой просто семейство гипербол $pv = \text{const}\,(=RT)$ с асимптотами $v=0$ и $p=0$; ни максимумов, ни минимумов они не имеют. Гораздо сложнее выглядят «изотермы ван дер Ваальса» (14), к рассмотрению которых мы и перейдем.

Найдем из (14) зависимость $p=p\,(v)$ давления от объема:

$$p = \frac{RT}{v-b} - \frac{a}{v^2}. \tag{7.1.15}$$

Из (15) вытекает

$$p' = \frac{dp}{dv} = -\frac{RT}{(v-b)^2} + \frac{2a}{v^3} = \\ = \frac{1}{(v-b)^2}\left[\frac{2a\,(v-b)^2}{v^3} - RT\right], \tag{7.1.16}$$

в силу чего возможные максимумы и минимумы давления p (при данной температуре T) определяются условием

$$\frac{2a\,(v-b)^2}{v^3} - RT = 0, \tag{7.1.17}$$

равносильным условию $p'=0$. Однако (17) представляет собой, к сожалению, кубическое уравнение относительно неизвестного v, решение которого вовсе не просто.

Не имея возможности решить уравнение (17), исследуем его качественно. Рассмотрим функцию $f\,(v) = \frac{2a\,(v-b)^2}{v^3} - RT$ и выясним характер ее изменения, найдем ее максимумы и минимумы. Для этого продифференцируем функцию f:

$$\frac{df}{dv} = 2a\left[\frac{2\,(v-b)}{v^3} - \frac{3\,(v-b)^2}{v^4}\right] = \\ = -\frac{2a\,(v-b)\,(v-3b)}{v^4}. \tag{7.1.18}$$

Из физических соображений (см. (14)) следует, что всегда $v > b$ (абсолютная температура T не может быть отрицательной!), поэтому (18) может обращаться в нуль лишь при $v=3b$. Кроме того, при $b < v < 3b$ правая часть (18) положительна, а при $v > 3b$ она отрицательна, поэтому значение $v=3b$ отвечает максимуму функции $f\,(v)$, наибольшему ее значению $f\,(v) = \frac{8a}{27b} - RT$. Далее надо отдельно рассмотреть три случая, реализующиеся при разных значениях температуры T.

а) $\frac{8a}{27b} - RT < 0$, т. е. $T > \frac{8a}{27bR}$. В этом случае $f\,(v) < 0$ при всех v, а, значит, $p' < 0$ при всех v; поэтому функция $p=p\,(v)$ все время убывает (см. две верхние из изображенных на рис. 9 изотерм ван дер Ваальса для углекислоты CO_2; температуры справа указаны по шкале Цельсия). Соответствующие кривые $p=p\,(v)$ по характеру напоминают гиперболы; асимптотами их служат прямые $p=0$ и $v=b$.

б) $\frac{8a}{27b} - RT > 0$, т. е. $T < \frac{8a}{27bR}$ (см. на рис. 9 две нижние кривые). Поскольку при $v=b$

[2] Ваальс Ян Дидерик ван дер (1837—1923) — голландский физик.

функция $f(v)$ отрицательна (нам не важно, что это значение $v=b$ не отвечает никакому реальному состоянию вещества!), а при $v=3b$ мы имеем $f(v)=\frac{8a}{27b}-RT>0$, то между b и $3b$ существует такое значение v_0 переменной v, что $f(v_0)=0$, а значит, и $p'(v_0)=0$; это значение (ранее которого производная p' отрицательна, а после положительна) отвечает минимуму функций $p(v)$. С другой стороны, при $v\to\infty$ в силу (15) $p\to 0$, а, значит, где-то при $v=v_1>3b$ функция $p=p(v)$ должна достигнуть максимума. Асимптоты кривой $p=p(v)$ и здесь остаются теми же, что и в случае а).

в) $\frac{8a}{27b}-RT=0$, т. е. $T=\frac{8a}{27bR}$. Так как наибольшее возможное значение функции $f(v)$ здесь равно нулю, то при всех других значениях v эта функция отрицательна, а тогда отрицательна и производная $p'=\frac{f(v)}{(v-b)^2}$. Таким образом, функция $p=p(v)$, подобно случаю а), все время убывает; при $v=3b$ она имеет перегиб (ибо $p'(3b)=0$).

В случае а), очевидно, каждому значению p отвечает единственное значение объема v, что соответствует возможности *при* $T>\frac{8a}{27bR}$ единственного физического состояния вещества — *газообразного*. Напротив, при $T<\frac{8a}{27bR}$ данному давлению p могут отвечать, как видно из рис. 9, три разных значения объема v, так что здесь вещество может существовать в разных состояниях. Состояние с наименьшим объемом $v=v_ж$ (т. е. с наибольшей плотностью) — это *жидкость*. Состояние с наибольшим объемом $v=v_г$ (при тех же давлении и температуре) представляет собой *газ*. Наконец, среднее состояние вещества оказывается неустойчивым. Если поместить вещество в сосуд определенного объема v (промежуточного между $v_ж$ и $v_г$) и нагреть до рассматриваемой температуры T ($T<T_к$ — см. ниже), то часть вещества будет находиться в газообразном, а другая — в жидком состоянии. Состояние со средним объемом v существовать не будет.

Пограничный случай $T=\frac{8a}{27bR}$ играет особо важную роль: температура T называется *критической* температурой газа (ее обычно обозначают через $T_к$). Точка перегиба K соответствующей изотермы отвечает следующим значениям объема и давления: $v=3b\,(=v_к)$ и $p=\frac{a}{27b^2}\,(=p_к)$; эти объем и давление также называют *критическими*. При $T=T_к$, если $p>p_к$ (или $v<v_к$), вещество находится в газообразном состоянии, а если $p<p_к$ (при $v>v_к$), оно уже становится жидким; таким образом, $p_к$ есть наибольшее давление, при котором вещество еще может существовать в виде насыщенного пара. Заметим еще, что точке K критической изотермы отвечает значение $p_кv_к=\frac{a}{9b}=\frac{3}{8}RT_к$, которое в этом случае составляет всего лишь 3/8 той величины, которую дает уравнение Клапейрона (13).

Более полное обсуждение уравнения ван дер Ваальса, его следствий и дальнейших уточнений можно найти в учебниках физики.

Обратимся теперь к вопросу о максимумах и минимумах функции двух переменных.

Определение $dy=y'dx$ дифференциала функции $y=f(x)$ (см. § 4.1) показывает, что точки (гладкого) максимума или минимума функции характеризуются тем, что в них дифференциал обращается в нуль; это тесно связано с геометрическим смыслом дифференциала (см. рис. 4.1.2) и с тем фактом, что касательная к графику функции в точке ее максимума или минимума должна быть горизонтальна. Столь же просто видеть, что также и в точке максимума или минимума функции $z=F(x, y)$ двух переменных дифференциал $dz=\frac{\partial F}{\partial x}dx+\frac{\partial F}{\partial y}dy$ функции F должен обращаться в нуль; иными словами, должны иметь место равенства $\frac{\partial F}{\partial x}=\frac{\partial F}{\partial y}=0$. Это связано как с тем, что касательная плоскость[3] к поверхности $z=F(x, y)$ в точке максимума или минимума должна быть горизонтальна (рис. 10, *а*), так и с тем, что в точке (x_0, y_0) максимума функции $z=F(x, y)$ достигают максимума и функции $f(x)=F(x, y_0)$ и $g(y)=F(x_0, y)$ одной переменной (x или y), и, значит, $f'(x)\left(=\frac{\partial F}{\partial x}\right)$ и $g'(y)\left(=\frac{\partial F}{\partial y}\right)$ должны равняться нулю. Однако надо иметь в виду, что если равенство $f'(x)=0$ почти всегда означает максимум или минимум функции f (ибо случаи обращения в нуль второй производной являются исключи-

[3] *Касательная плоскость* в точке $M=M(x_0, y_0)$ к поверхности П с уравнением $z=F(x, y)$ содержит касательные к линиям $y=y_0$, $z=F(x, y_0)=f(x)$ и $x=x_0$, $z=F(x_0, y)=g(y)$ (к линиям $y=\text{const}$ и $x=\text{const}$); из всех проходящих через M плоскостей она теснее всего примыкает к поверхности П.

тельными), то условия $\frac{\partial F(x_0, y_0)}{\partial x} = \frac{\partial F(x_0, y_0)}{\partial y} = 0$ (или $dz=0$) могут означать и то, что одна из фигурирующих выше функций $f(x)$ и $g(x)$

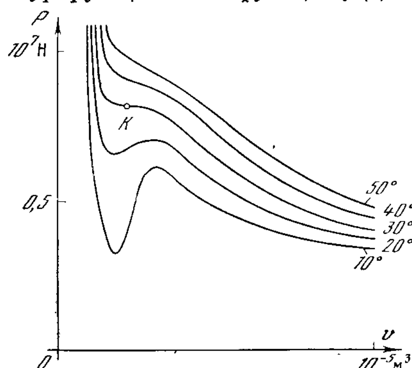


Рис. 7.1.9

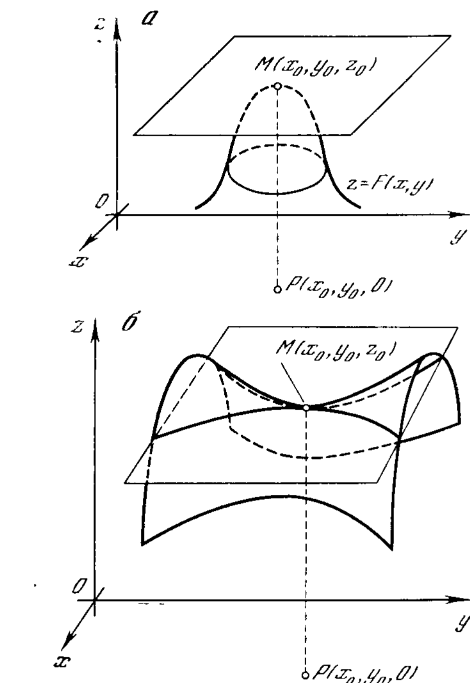


Рис. 7.1.10.

достигает в рассматриваемой точке (x_0, y_0) максимума, а вторая — минимума. В этом случае точка (x_0, y_0) соответствует «седлу» (или «перевалу») поверхности $z=F(x, y)$ и в ней функция F не достигает ни максимума, ни минимума (рис. 10, *б*).

Упражнения

7.1.1. Из прямоугольного жестяного листа со сторонами a и b делают ящик, вырезая равные квадраты по углам. Какова должна быть сторона у вырезанных квадратов, чтобы ящик имел максимальный объем?

7.1.2. В остроугольный треугольник с основанием a и высотой h впишите прямоугольник, две вершины которого принадлежат основанию, а две другие — боковым сторонам треугольника, причем так, чтобы этот прямоугольник имел наибольшую возможную площадь.

7.1.3. Определите наибольшую возможную площадь прямоугольника, вписанного в круг данного радиуса R.

7.1.4. При каком радиусе основания и при какой высоте закрытая цилиндрическая банка данного объема V будет иметь наименьшую площадь (полной) поверхности?

7.1.5. Два тела движутся по сторонам прямого угла по направлению к вершине с постоянными скоростями v_1 и v_2 (м/с). Через сколько секунд после начала движения расстояние между телами будет наименьшим, если в начале движения первое находилось на расстоянии a м от вершины угла, а второе — на расстоянии b м?

7.1.6. Докажите, что произведение двух положительных чисел, сумма которых постоянна, будет наибольшим, если эти числа равны.

7.1.7. Прямой l плоскость разделена на две части (среды I и II). Тело движется в среде I со скоростью v_1, а в среде II — со скоростью v_2. По какому пути должно двигаться тело, для того чтобы возможно скорее попасть из данной точки A среды I в данную точку B среды II? (Эта задача ставит вопрос о з а к о н е п р е л о м л е н и я с в е т а при переходе из среды I в среду II, в которых свет распространяется с разными скоростями: ведь известно, что траектория светового пучка между точками A и B всегда такова, что свет проходит путь от A до B в кратчайшее время (*принцип Ферма*, см. с. 109).)

§ 2. Негладкие максимумы и минимумы. Изломы и разрывы. Левая и правая производные функции

До сих пор мы говорили, что максимумы и минимумы функции достигаются при тех значениях x, при которых первая производная обращается в нуль. Однако иногда максимумы (и минимумы) отвечают и значениям аргумента, н е о б р а щ а ю щ и м (первую) производную в нуль.

Для примера вернемся снова к рассмотренной в предыдущем параграфе задаче о последовательно включенных в сеть с данной ЭДС сопротивлениях r

и R (см. рис. 1.3) и спросим себя, при каком сопротивлении R энергия W, выделяемая на сопротивлении r будет максимально возможной.

В точности аналогично (1.6) имеем

$$W = i^2 r = \frac{u^2 r}{(r+R)^2}. \qquad (7.2.1)$$

Напомним, что в (1) мы считаем u и r известными, а сопротивление R — переменным и неизвестным. Из (1) сразу следует

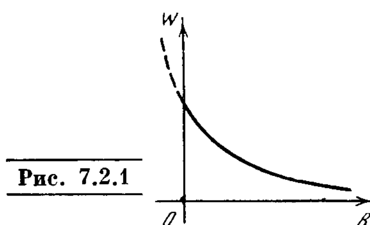


Рис. 7.2.1

$$\frac{dW}{dR} = -\frac{2u^2 r}{(r+R)^3};$$

таким образом, условие $\frac{dw}{dR} = 0$ обращается здесь в равенство

$$\frac{u^2 r}{(r+R)^3} = 0,$$

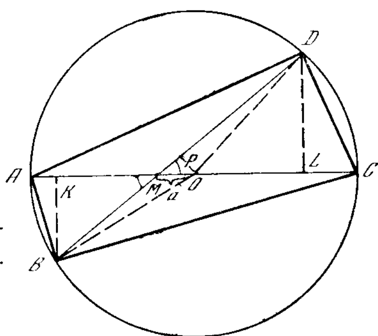


Рис. 7.2.2

не выполняющееся ни при каком R.

Значит ли это, что мощность может расти неограниченно, что задача о максимальной мощности не имеет решения? Ведь из физического смысла задачи ясно, что мощность будет наибольшей при $R = 0$ $\left(\text{в этом случае } W = \frac{u^2}{r}\right)$. Почему же из уравнения $\frac{dW}{dR} = 0$ мы не получили значения $R = 0$?

Для того чтобы разобраться в этом, рассмотрим график зависимости $W = W(R)$ (рис. 1).

Из графика видно, что если бы R могло принимать и отрицательные значения, то при $R=0$ максимума бы не было. Однако из физического содержания задачи следует, что отрицательные R невозможны: всегда $R \geqslant 0$. Таким образом, величина W имеет максимум при $R=0$ из-за того, что ограничен промежуток изменения аргумента. Этот вывод следует запомнить: если промежуток изменения аргумента ограничен, то при исследовании функции на максимум и минимум надо специально рассмотреть *граничные* значения аргумента.

В качестве еще одного примера обратимся к уже рассматривавшемуся в § 2.6 вопросу о возможных значениях теплоемкости; только теперь будем решать этот вопрос не для алмаза, а для воды. Количество тепла (в джоулях), необходимое для нагревания 1 кг воды (при атмосферном давлении) от 0 до T °С, приближенно равно $Q(T) = 4186{,}68\,T + 8373{,}36 \cdot 10^{-5}\,T^2 + 1256 \cdot 10^{-6}\,T^3$, откуда следует, что теплоемкость воды $c = c(T)$ при температуре T равна

$$c(T) = \frac{dQ}{dT} = 4186{,}68 + 16746{,}72 \cdot 10^{-5} T + 3768 \cdot 10^{-6} T^2$$

(см. упр. 2.2.2 и его решение). Ясно, что при положительных T величина $c(T)$ с ростом T растет, а при уменьшении T падает. Означает ли это, что теплоемкость $c(T)$ воды может быть сколь угодно велика? Разумеется, нет: ведь вода может существовать лишь в интервале температур от 0 до 100° С и на левом конце этого интервала теплоемкость принимает наименьшее, а на правом — наибольшее возможное значение.

Рассмотрим, наконец, следующую геометрическую задачу (рис. 2): на диаметре AC окружности выбрана точка M; под каким углом α к диаметру следует провести через M хорду BD, чтобы (вписанный в окружность) четырехугольник $ABCD$ имел наибольшую возможную площадь?

Примем радиус окружности за единицу длины; расстояние OM точки M от центра окружности обозначим через a. Поскольку очевидно, что высоты BK и DL треугольников ABC и ADC рав-

ны соответственно $BM \sin\alpha$ и $DM \sin\alpha$, то

$$S_{ABCD} = S = S_{\triangle ABC} + S_{\triangle ADC} = \frac{1}{2} AC \cdot BM \sin\alpha + \frac{1}{2} AC \cdot DM \sin\alpha = \frac{1}{2} AC (BM + DM) \sin\alpha = \frac{1}{2} AC \cdot BD \sin\alpha.$$

Кроме того, расстояние OP от центра до хорды BD равно $OM \sin\alpha = a \sin\alpha$; поэтому $BP = PD = \sqrt{1 - a^2 \sin^2\alpha}$ (ибо радиус $OB = OD$ окружности мы приняли за единицу), и, значит, $BD = 2\sqrt{1 - a^2 \sin^2\alpha}$, т. е.

$$S = \frac{1}{2} AC \cdot BD \sin\alpha = 2 \sin\alpha \sqrt{1 - a^2 \sin^2\alpha}. \quad \mathbf{(7.2.2)}$$

По правилам дифференцирования отсюда следует

$$\frac{dS}{d\alpha} = \frac{2\cos\alpha\,(1 - 2a^2 \sin^2\alpha)}{\sqrt{1 - a^2 \sin^2\alpha}}. \quad (7.2.3)$$

Из (3) видно, что обращение производной в нуль равносильно либо равенству $\cos\alpha = 0$, либо равенству $\sin\alpha = 1/(a\sqrt{2})$. Первый случай соответствует значению $\alpha = 90°$, т. е. хорда $BD \perp AC$; второй возможен не при всех значениях a. Очевидно, что лишь при $a > \sqrt{2}/2$ ($\approx 0{,}71$) мы получим $\sin\alpha < 1$ и сможем найти соответствующее значение угла $\alpha < 90°$.

Здесь мы сталкиваемся с ситуацией, когда задача, в условие которой входит параметр, имеет решения качественно разного типа в зависимости от численного значения параметра. Если $a \leqslant \sqrt{2}/2$, то функция $S = S(x)$, при крайних значениях аргумента $\alpha = 0$ и $\alpha = 180°$ обращающаяся, очевидно, в нуль, может иметь максимум лишь при $\alpha = 90°$, а так как максимум она, безусловно, имеет (ср. с. 187—188), то значение $\alpha = 90°$ наверняка отвечает максимальной величине площади $S(\alpha)$ в силу (2) равной[1] $2\sqrt{1 - a^2}$. Но при $a > \sqrt{2}/2$ максимум $S(\alpha)$ может достигаться при трех значениях α: при $\alpha = 90°$ и при двух значениях $\alpha = 90° \pm \beta$, таких, что $\sin\alpha = 1/(a\sqrt{2})$. Здесь, напротив, максимум функции $S(\alpha)$ реализуется при двух последних значениях α, которым отвечают две симметричные относительно диаметра AC хорды BD; соответствующее значение (2) площади S обращается в $1/a$, значению же $\alpha = 90°$ отвечает (локальный) минимум функции $S(\alpha)$. (Заметим еще, что если искать не максимум функции $S = S(\alpha)$, а максимум более простой функции $F(x) = x - a^2x^2$, где $F = [S(x)]^2/4$ и $x = \sin^2\alpha$, то значение $\alpha = 90°$ будет отвечать граничному максимуму функции $F(x)$, где $0 \leqslant x \leqslant 1$, а равенство $\sin\alpha = 1/(a\sqrt{2})$ — гладкому максимуму F.)

Чтобы разобраться полностью в задаче, постройте самостоятельно графики функции $S = S(\alpha)$ (см. (2)) при трех (или четырех) фиксированных значениях a: $a = 0{,}9$; $\sqrt{2}/2$ и $0{,}5$ (или даже при $a = 0$).

На рис. 3 изображен график функции $\alpha = \alpha(a)$, где угол α таков, что площадь четырехугольника $ABCD$ на рис. 2 максимальна. Заметьте, что при $a > \sqrt{2}/2$ задача о выборе хорды BD имеет два симметричных относительно AC решения, в то время как при $a \leqslant \sqrt{2}/2$ решение будет только одно.

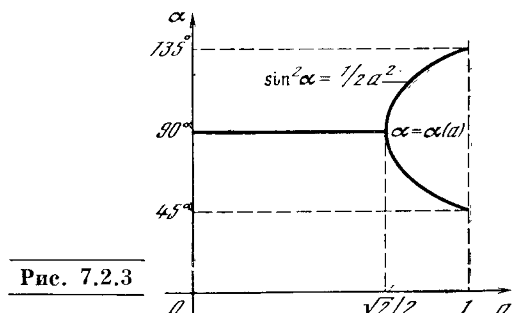


Рис. 7.2.3

Вернемся к общему вопросу о граничных и гладких максимумах и минимумах функциий.

Если максимум (минимум) функции $y = f(x)$ достигается на краю x_0 промежутка изменения аргумента (такие максимумы и минимумы иногда называют *граничными*), то первый член ряда

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \ldots$$

может и не выпадать. Поэтому если мы несколько отступили от (граничного!) значения x_0, которому отвечает максимум, то ошибка в определении y может быть и велика: ведь она будет пропорциональна $(x - x_0)$, а не $(x - x_0)^2$, как ранее. Поэтому здесь даже

[1] Таким образом, график зависимости $S_{\max} = S(a)$ в нашем случае изображается (в плоскости с координатами a, S) линией, которая состоит из дуги эллипса $a^2 + \left(\frac{S}{2}\right)^2 = 1$ $\left(\text{при } 0 \leqslant a \leqslant \frac{\sqrt{2}}{2}\right)$ и дуги гиперболы $S = \frac{1}{a}$ $\left(\text{при } \frac{\sqrt{2}}{2} \leqslant a \leqslant 1\right)$.

малая ошибка в определении значения аргумента, отвечающего максимуму функции, может привести к ощутимой погрешности в значении функции.

В рассмотренных примерах функция $f(x)$ существовала и при $x < x_0$, но физического (или геометрического) смысла не имела. Однако может случиться и так, что функция $f(x)$ при некоторых значениях аргумента вовсе не существует. Не существует, например, корень четной степени (скажем, квадратный) из отрицательного числа, и здесь значения аргумента, обращающие подкоренное выражение в нуль, являются граничными. При исследовании на максимум или минимум такие значения аргумента должны быть рассмотрены отдельно. Так, если $y = a - \sqrt{b - x}$, то $y' = 1/2\sqrt{b - x}$, и, хотя y' нигде не обращается в нуль, функция y имеет максимум. Этот максимум достигается при $x = b$: в самом деле, значение $x = b$ обращает подкоренное выражение в нуль, больше же b переменное x быть не может, так как иначе подкоренное выражение станет отрицательным. При $x = b$ мы имеем $y = a$; при всех же других значениях x величина y получается вычитанием из a положительного числа[2] $\sqrt{b - x}$ и поэтому будет меньше a.

Максимум (или минимум) может достигаться и во внутренних точках, в которых производная не обращается в нуль, если кривая в этой точке имеет *излом, угловую точку*. Такие точки встречаются, в частности, когда кривая состоит из двух частей, описываемых разными формулами при $x < x_0$ и при $x > x_0$. Приведем пример физической задачи такого рода.

Пусть на электроплитке постоянной мощности нагревается чайник; требуется определить тот момент, когда количество теплоты, запасенное чайником, будет наибольшим. Для простоты условимся считать, что коэффициент полезного действия плитки равен 100%, т. е. что все тепло она отдает чайнику. Пусть мы поставили чайник на плитку в момент времени $t = 0$ и пусть в этот момент чайник обладал q джоулями тепла[3]. По определению единицы количества тепла (джоуль) общее количество тепла Q (в джоулях), выделенное электроплиткой, равно i^2Rt, где i — сила тока (в амперах), R — сопротивление плитки (в омах), t — время (в секундах). Поэтому количество тепла в чайнике в момент t равно

$$Q = q + i^2Rt.$$

В некоторый момент времени $t = t_0$, когда количество накопленного чайником тепла равно $Q_0 = q + i^2Rt_0$, чайник закипает. В этот момент вода начинает превращаться в пар (выкипать)[4]. На образование 1 г пара идет примерно 2256,7 Дж тепла. Обозначая через dm массу воды, выкипевшей за время dt, получим

$$dm \approx (i^2R/2256,7)\,dt \approx 45 \cdot 10^{-5} i^2R dt.$$

Следовательно, за 1 с выкипает $\frac{dm}{dt} \approx \approx 45 \cdot 10^{-5} iR^2$ г воды; на это затрачивается

$$\frac{dQ_1}{dt} \approx 418,68 \frac{dm}{dt} \approx 418,68 \cdot 45 \cdot 10^{-5} iR^2 \approx \approx 0,1884 i^2R \text{ Дж/с}$$

тепла, поскольку 1 г воды при температуре кипения содержит примерно 418,68 Дж тепла. Поэтому к моменту времени t, где $t > t_0$, на превращение выкипевшей воды в пар будет затрачено

$$Q_1 \approx 0,1884 i^2R (t - t_0)$$

тепла (в Дж).

Мы видим, что количество тепла Q (в Дж) в чайнике выражается двумя разными формулами: при $t < t_0$ (т. е. до начала кипения) оно равно

$$Q = q + i^2Rt,$$

а при $t > t_0$ (после того как чайник закипел) — равно

$$Q \approx q + i^2Rt_0 - 0,1884 i^2R (t - t_0) = = q + i^2R (1,1884 t_0 - 0,1884 t).$$

---

[2] Мы считаем, что под $\sqrt{b-x}$ понимается положительное (арифметическое) значение корня.

[3] За нуль принимается тепловая энергия воды при 0° С.

[4] Парообразование происходит и при температуре, меньшей 100° С, т. е. до начала кипения воды, но мы им пренебрегаем.

График зависимости $Q=Q(t)$ изображен на рис. 4, *а*. Из этого рисунка видно, что функция $Q(t)$ достигает максимума при $t=t_0$, хотя производная $Q'(t)$ при этом значении t в нуль не обращается, т. е. касательная к графику в точке $(t_0, Q(t_0))$ не горизонтальна (функция $Q(t)$ при $t=t_0$ вообще не имеет производной, а график этой функции в соответствующей точке не имеет касательной).

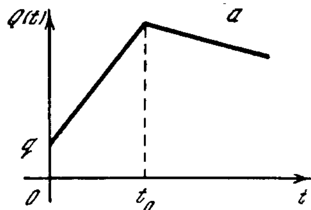
Производная функции $Q=Q(t)$ при $t=t_0$ терпит р а з р ы в. Действительно, если ограничиться лишь значениями $t<t_0$, то нам придется считать, что $Q'(t)=i^2R$; в точках же, где $t>t_0$, производная $Q'(t)\approx-0{,}1884i^2R$. График производной $\frac{dQ}{dt}=Q'(t)$ изображен на рис. 4, *б*.

Итак, максимум функции $y=f(x)$ может достигаться в у г л о в о й точке, где производная разрывна и изображающая график линия образует угол.

Из рис. 5 также видно, что минимум (или максимум) может быть достигнут при тех значениях аргумента x_0, где производная терпит бесконечный разрыв. (На рис. 5 изображен график функции $y=x^{2/3}=\sqrt[3]{x^2}$; график обратной ей функции $y=x^{3/2}$, или $y^2=x^3$, представляет собой п о л у к у б и ч е с к у ю п а р а б о л у; см. § 1.5 и, в частности, рис. 1.5.7.) Точка кривой такого типа, как та, в которой достигается минимум функции $y=x^{2/3}$, называется *острием*. График производной функции $y=x^{2/3}$ изображен на рис. 6. Здесь, как и в случае обычного минимума, $y'<0$ при $x<x_0$ (на рис. 6 $x_0=0$) и при приближении к $x=x_0$ слева функция убывает. При $x>x_0$ имеем $y'>0$: при увеличении x, после того как значение $x=x_0$ пройдено, функция возрастает. Однако при $x=x_0$ производная теряет смысл. Она положительна и сколь угодно велика при x, близком к x_0 и большем x_0, и отрицательна и сколь угодно велика по абсолютной величине при x, близком к x_0 и меньшем x_0.

Максимумы и минимумы, достигаемые при тех значениях аргумента, когда производная терпит разрыв, иногда называют *острыми* максимумами и минимумами; *острые* и *граничные* (достигаемые на границе области определения функции) максимумы и минимумы вместе можно назвать *негладкими*.

В связи с рассмотрением особых точек на кривых и в первую очередь



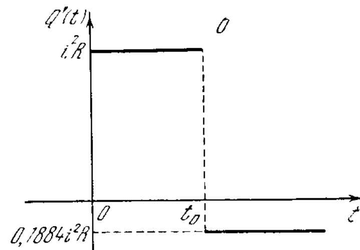


Рис. 7.2.4

точек излома (см. рис. 4, *а*) можно уточнить рассуждения, которые привели нас к понятию производной. В гл. 2

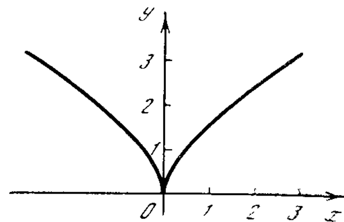


Рис. 7.2.5

книги, не оговаривая этого особо, мы всюду рассматривали лишь *гладкие* кривые. При этом производная $y'(t)$,

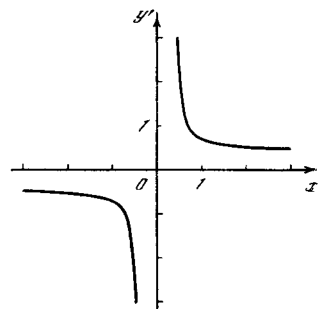


Рис. 7.2.6

взятая в точке t, определялась как предел отношения

$$\frac{y(t_2)-y(t_1)}{t_2-t_1} \qquad (7.2.4)$$

при стремлении t_2 и t_1 к t (здесь разность t_2-t_1, очевидно, стремится к нулю). Специально подчеркивалось, что этот предел не зависит от того,

как выбраны значения t_2 и t_1: они **могут** быть оба больше t, или оба **меньше** t, или одно из них больше, **а** другое меньше t, или одно из них **равно** t, а другое больше или меньше t. **Так,** при положительном Δt дробь $\frac{y(t+\Delta t)-y(t)}{\Delta t}$ соответствует значениям в (4) $t_1=t$ и $t_2=t+\Delta t>t$, а дробь $\frac{y(t)-y(t-\Delta t)}{\Delta t}$ — значениям $t_1=t-\Delta t<t$ и $t_2=t$ [5].

Если функция $y=y(x)$ — гладкая, то во всех этих случаях получается

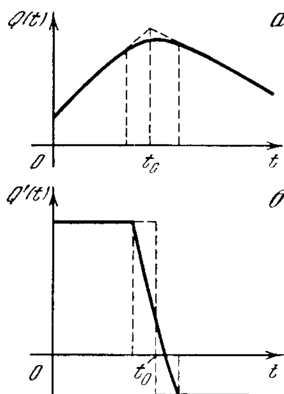


Рис. 7.2.7

один и тот же предел, равный **производной** в данной точке. Однако в случае кривой с изломом (ср. рис. 4) положение дела меняется. Если обозначить значение t, при котором имеет место излом, через t_0, то дробь $\frac{y(t_0+\Delta t)-y(t_0)}{\Delta t}$ стремится при п о л о ж и т е л ь н о м и стремящемся к нулю Δt к определенному значению (в примере на с. 196 — к значению $-0{,}1884i^2R$), которое называют *правой производной* функции $y(x)$ в точке $t=t_0$. Образовав дробь $\frac{y(t_0)-y(t_0-\Delta t)}{\Delta t}$, **мы** получим при п о л о ж и т е л ь н о м и стремящемся к нулю Δt другой предел, называемый *левой производной* функции (в упомянутом примере левая производная равна i^2R).

[5] Для г л а д к и х кривых мы вычисляли производную тоже как предел отношения
$$\left[y\left(t+\frac{\Delta t}{2}\right)-y\left(t-\frac{\Delta t}{2}\right)\right]\Big/\Delta t$$
при $\Delta t\to 0$ (см. текст, напечатанный на с. 155 мелким шрифтом); здесь $t_1=t-\frac{\Delta t}{2}<t$, $t_2=t+\frac{\Delta t}{2}>t$.

Взяв t_2 и t_1 с разных сторон от t_0, можно получить в пределе при $t_2\to t_0$, $t_1\to t_0$ р а з н ы е значения отношения (4): нетрудно видеть, что если A есть угловая точка кривой или острие, то хорда BC, где B и C лежат по разные стороны от A, может иметь разные направления (представьте себе такую хорду на рис. 4, *а* и 5). Поэтому предел этой хорды при $B, C\to A$ тоже может быть разным: он зависит от того, как именно стремятся точки B и C к точке A. Таким образом, в самой точке излома графика функции (в точке «излома функции», как иногда в таких случаях говорят) производная не имеет определенного значения, но здесь можно определить производную слева и производную справа.

В гл. 2, впервые изучая производные, мы намеренно для упрощения изложения не отмечали каждый раз, что определенное значение производной, не зависящее от способа стремления x к x_0 (например, слева или справа), существует лишь для точек, в которых график функции — г л а д к а я кривая. Как видно из рис. 4, *б*, в точке, где кривая $y(t)$ имеет и з л о м, график производной $y'(t)$ терпит р а з р ы в. Если излом на кривой $y(t)$ заменить дугой малого радиуса, касающейся кривой слева и справа (как говорят чертежники, сделать *сопряжение*), то полученной гладкой кривой будет отвечать непрерывно меняющаяся производная; однако на том участке изменения t, где кривая $y(t)$ заменена дугой, кривая $y'(t)$ круто меняется (ср. рис.7 с рис. 4).

Если график функции $y(t)$ имеет разрыв в точке t_0 (рис. 8, *а*), то условно можно сказать, что в точке t_0 производная $y'(t)$ б е с к о н е ч н а (хотя на самом деле в этой точке, конечно, вовсе нет производной). В самом деле, если разрыв заменить гладким изменением y от y_1 до y_2 на малом отрезке от $t_0-\varepsilon$ до $t_0+\varepsilon$, то на этом отрезке производная равна $\frac{y_2-y_1}{2\varepsilon}$, т. е. очень велика, и тем больше, чем меньше ε (рис. 8, *б*).

А как обстоит дело с и н т е г р а л о м $\int_a^b y(t)\,dt$, если функция $y(t)$ не гладкая? Если функция имеет угловые

точки, то при вычислении площади, ограниченной кривой, никаких новых вопросов не возникает. В § 3.2 мы рассматривали определенный интеграл (площадь) как сумму площадей прямоугольных полосок вида $y(t_n)(t_{n+1}-t_n)$ или $y(t_{n+1})(t_{n+1}-t_n)$. В пределе, при уменьшении величин интервалов, т. е. разностей $(t_{n+1}-t_n)$, становится безразлично, брать ли в сумме каждый раз $y(t_n)$ или $y(t_{n+1})$, причем так будет обстоять дело как в случае гладкой кривой $y(t)$, так и в случае кривой $y(t)$, имеющей излом.

Если кривая $y(t)$ терпит разрыв в точке $t=t_0$, но остается ограниченной, то для интервала, внутри которого находится разрыв $(t_n < t_0 < t_{n+1})$, величины $y(t_n)$ и $y(t_{n+1})$ остаются различными, как бы мы ни сближали t_n и t_{n+1}. Таким образом, в выражении для интегральной суммы величина одного из слагаемых в этом случае существенно зависит от того, как берется сумма: по формуле типа (3.2.1) или типа (3.2.2). Однако при стремлении к нулю величины интервала $t_{n+1}-t_n$ само это слагаемое стремится к нулю, поэтому предел суммы, т. е. интеграл, имеет вполне определенное значение (не зависящее от способа вычисления суммы) и в том случае, когда подынтегральная функция претерпевает в области интегрирования разрыв. Здесь, разумеется, мы считаем, что значения функции слева и справа от разрыва, а значит, и величины разрыва являются конечными, т. е. что функция не ведет себя, скажем, как функция $y=\frac{1}{x}$ в окрестности значения $x=0$.

Сохраняется и соотношение между интегралом и производной. Обратимся снова к рис. 4. Если обозначить функцию $Q'(t)$, график которой изображен на рис. 4, *б*, через $f(t)$, то исходная функция $Q(t)$ (см. рис. 4, *а*) будет представлять собой интеграл $Q(t)=\int f(t)\,dt$. На этом примере мы видим, что разрыв подынтегральной функции $f(t)=Q'(t)$ приводит к излому графика интеграла $Q(t)$ этой функции. Определенный интеграл функции с разрывом такого рода может быть найден с помощью неопределенного интеграла по общему правилу

$$\int_a^b f(t)\,dt = Q(b)-Q(a).$$

Можно пойти и дальше: исходя из рис. 8, можно сказать, что для функции, стремящейся к бесконечности на интервале, стремящемся к нулю (рис. 8, *б*) [6], интеграл является **р а з р ы в н о й** функцией (рис. 8, *а*). Однако при этом надо уточнить характер стремления функции к бесконечности и интервала — к нулю; здесь мы на этом останавливаться не будем. Такого рода примеры (более полное рассмотрение которых требует еще некоторых уточнений) приводят к понятию δ-функции, которому посвящена гл. 16 третьей части книги.

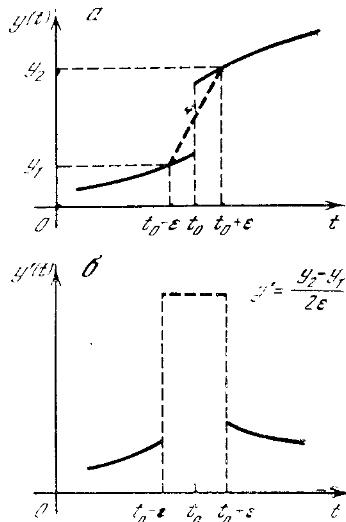


Рис. 7.2.8

Итак, окончательно мы приходим к следующей схеме решения задач на отыскание максимумов и минимумов функций. Прежде всего надо найти производную заданной функции и определить так называемые стационарные точки функции (т. е. точки, в которых функция изменяется медленнее всего) — точки обращения в нуль производной. Но наряду с этими точками в разряд

[6] Выражения «интервал, стремящийся к нулю», «функция, стремящаяся к бесконечности», указывают, что по существу здесь речь идет не о функции, а **о с е м е й с т в е** функций, характеризующемся тем, что при переходе от одной функции семейства к следующей «интервал роста» функции становится все уже, а «степень роста» — все больше. Подробнее эта ситуация будет освещена в гл. 16.

«подозрительных» (т. е. тех, в которых может достигаться максимум или минимум) попадают также граничные точки области определения функции, точки излома функции, точки, в которых производная обращается в бесконечность, меняя свой знак («острия»). Далее надо проверить характер изменения функции во всех этих точках. Если в точке, где $f'(x)=0$, график функции представляет собой гладкую кривую, то информацию о типе точки (максимум; минимум; ни максимум, ни минимум) можно получить, определив вторую производную функции (или старшие производные, если первая и вторая производные в этой точке обращаются в нуль одновременно). В других же случаях представление о поведении функции в рассматриваемой точке могут дать односторонние (т. е. левая и правая) производные (ясно, что в граничных точках области определения функции существуют лишь односторонние ее производные). Наконец, если нас интересует *абсолютный максимум* или *минимум* функции на рассматриваемом интервале ее изменения, т. е. самый большой из всех (локальных) максимумов или самый малый из всех минимумов, то надо будет еще сравнить значения функции во всех точках ее локальных максимумов (или минимумов).

Упражнения

7.2.1. Найдите наименьшее значение функции $y=x^2-2x+3$, определенной на интервале $2 \leqslant x \leqslant 10$.

7.2.2. На берегу реки сидит рыбак P; по реке, на расстоянии h от берега, на котором сидит P, плывет со скоростью v пароход Π длины l, нос которого проходит строго против рыбака в момент времени $t=t_0$. (Шириной парохода мы пренебрегаем, путь парохода считаем точно прямолинейным.) Расстоянием от P до Π естественно назвать расстояние от P до б л и ж а й ш е й к P точки Π. Как меняется это расстояние D с течением времени t? Когда достигает минимума функция $D=D(t)$? Нарисуйте график функции $D=D(t)$ для случая $h=300$ м, $l=60$ м, $v=5$ м/с.

7.2.3. Найдите «острые» максимумы функций: а) $y=(x-5)\sqrt[3]{x^2}$; б) $y=1-\sqrt[3]{x^2}$.

§ 3*. Выпуклые функции и алгебраические неравенства

В § 1.4 мы следующим образом определили *выпуклость кривой* $y=f(x)$ (или функции $y=f(x)$): функция $y=f(x)$ называется *выпуклой* на интервале $a \leqslant x \leqslant b$ изменения x (или кривая $y=f(x)$ *обращена выпуклостью вверх, выпукла*), если любая хорда AB этой кривой ($A=A(x_1, f(x_1))$, $B=B(x_2, f(x_2))$; здесь $a \leqslant x_1 < x_2 \leqslant b$) лежит н и ж е соответствующей дуги AB кривой $y=f(x)$ (рис. 1, *а*); напротив, если хорда AB лежит в ы ш е дуги AB линии $y=f(x)$, то функция f называется *вогнутой* (то кривая $y=f(x)$ *обращена выпуклостью вниз, вогнута*; см. рис. 1, *б*). Другими словами, функция $y=f(x)$ *выпукла* на интервале $a \leqslant x \leqslant b$, если точка N пересечения прямой $x=\text{const}$ ($a \leqslant x_1 < x < x_2 \leqslant b$) с хордой AB лежит н и ж е точки M пересечения той же прямой с графиком функции $y=f(x)$; напротив, если N лежит в ы ш е M, то функция f *вогнута*. Это определение уже связывает понятие выпуклости с неравенствами: ведь если точки M и N имеют координаты (x, y) и (x, Y), то выпуклость функции $y=f(x)$ равносильна неравенству $y>Y$ (а ее вогнутость — неравенству $y<Y$).

В § 2.7 мы указали простой признак выпуклости функции $y=f(x)$: он состоит в *отрицательности* ее *второй производной* $y''=\frac{d^2y}{dx^2}$. Так, например, из этого критерия выпуклости сразу вытекает, что выпуклы функции $y=\ln x$, $y=-x^2$, $y=-\frac{1}{x}$ (при $x>0$), поскольку их вторые производные, очевидно, равны соответственно: $y''=\left(\frac{1}{x}\right)'=-\frac{1}{x^2}$; $y''=(-2x)'=-2$; $y''=\left(\frac{1}{x^2}\right)'=-\frac{2}{x^3}$ (рис. 2; напомним, что функция $y=\ln x$ определена лишь при $x>0$).

Имеет место простая, но важная

Т е о р е м а 1. *Если $y=f(x)$ — выпуклая на отрезке от a до b функция, x_1 и x_2 — два значения аргумента этой функции, взятые внутри рассматриваемого отрезка (т. е. два произволь-*

ных числа, таких, что $a \leqslant x_1 < x_2 \leqslant b$), *то*

$$\frac{f(x_1)+f(x_2)}{2} < f\left(\frac{x_1+x_2}{2}\right). \qquad (7.3.1)$$

Доказательство. Пусть на рис. 3 $OA = x_1$, $OB = x_2$, в таком случае $AM = f(x_1)$, $BN = f(x_2)$. Далее, если S есть середина отрезка AB, то $OS = \frac{x_1+x_2}{2}$ и, следовательно, $SP = f\left(\frac{x_1+x_2}{2}\right)$. С другой стороны, так как средняя линия SQ трапеции $ABNM$ равна полусумме оснований AM и BN, то $SQ = \frac{f(x_1)+f(x_2)}{2}$. Но согласно определению выпуклой функции середина Q хорды MN расположена ниже точки P дуги MN; следовательно,

$$\frac{f(x_1)+f(x_2)}{2} < f\left(\frac{x_1+x_2}{2}\right),$$

что и требовалось доказать [1].

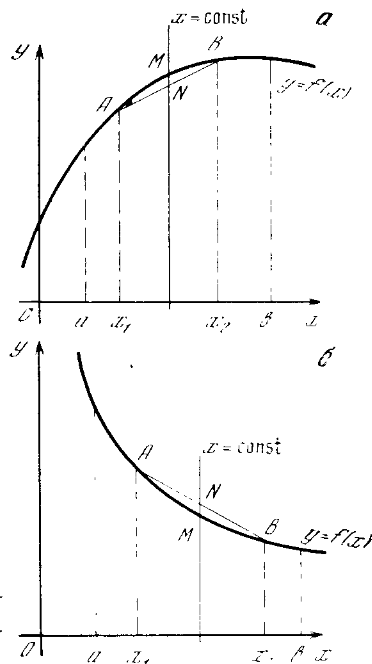


Рис. 7.3.1

Примеры.

а) $y = \ln x$. Из выпуклости этой функции следует, что

$$\frac{\ln x_1 + \ln x_2}{2} < \ln\frac{x_1+x_2}{2},$$

т. е. $\ln\sqrt{x_1x_2} < \ln\frac{x_1+x_2}{2}$,

или, наконец,

$$\sqrt{x_1x_2} < \frac{x_1+x_2}{2} \qquad (7.3.2)$$

— среднее геометрическое двух неравных положительных чисел меньше их среднего арифметического.

б) $y = -x^2$. Аналогично случаю а) здесь получаем

$$-\frac{x_1^2+x_2^2}{2} < -\left(\frac{x_1+x_2}{2}\right)^2,$$

или, в другой форме,

$$\frac{x_1^2+x_2^2}{2} > \left(\frac{x_1+x_2}{2}\right)^2, \qquad \sqrt{\frac{x_1^2+x_2^2}{2}} > \frac{x_1+x_2}{2}.$$

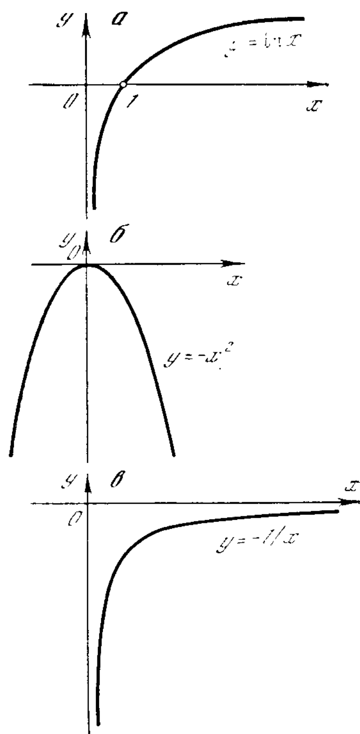


Рис. 7.3.2

Выражение $\sqrt{\frac{a_1^2+a_2^2+\ldots+a_k^2}{k}}$ — корень квадратный из среднего арифметического квадратов k чисел a_1, a_2, ..., a_k — называется *средним квадратичным* этих чисел. Таким образом, полученный результат можно сформу-

[1] Мы ограничиваемся при доказательстве теоремы 1 и дальнейших теорем случаем, когда $f(x_1)$ и $f(x_2)$ имеют одинаковые знаки. Предоставляем читателю самостоятельно рассмотреть случай разных знаков $f(x_1)$ и $f(x_2)$ (здесь вместо свойства средней линии трапеции придется применить следующую теорему: отрезок средней линии трапеции, заключенный между ее диагоналями, равен полуразности оснований трапеции).

лировать так: *среднее квадратичное двух* **неравных** *положительных чисел всегда* **больше** *их среднего арифметического.*

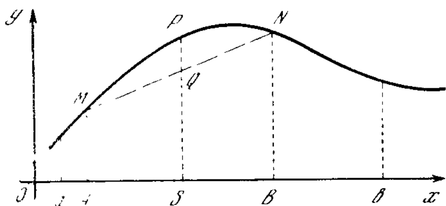


Рис. 7.3.3

в) $y=-\frac{1}{x}$. Из теоремы 1 следует

$$-\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)<-\frac{1}{(x_1+x_2)/2},$$

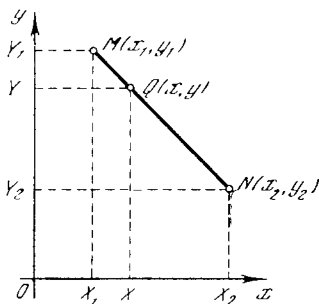


Рис. 7.3.4

или

$$\frac{1}{2}\left(\frac{1}{x_1}+\frac{1}{x_2}\right)>\frac{2}{x_1+x_2},$$

или, наконец,

$$\frac{2}{1/x_1+1/x_2}<\frac{x_1+x_2}{2}. \qquad (7.3.3)$$

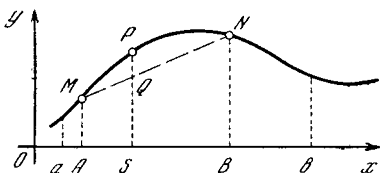


Рис. 7.3.5

Частное $1:\frac{1/a_1+1/a_2+\ \dots\ +1/a_k}{k}$ $\left(=\frac{k}{1/a_1+1/a_2+\ \dots\ +1/a_k}\right)$, обратное среднему арифметическому чисел $\frac{1}{a_1}$, $\frac{1}{a_2}$, ..., $\frac{1}{a_k}$, обратных k положительным числам $a_1, a_2, \dots, a_k$ (т. е. обратное числу $\frac{1}{k}\left(\frac{1}{a_1}+\frac{1}{a_2}+\dots+\frac{1}{a_k}\right)$) называется *средним гармоническим* чисел $a_1, a_2, \dots, a_k$. Таким образом, неравенство (3) утверждает, что *среднее гармоническое двух неравных положительных чисел меньше их среднего арифметического.*

Теорема 1 может быть обобщена следующим образом:

Теорема 2. *Если функция $y=f(x)$ выпукла в интервале от a до b, x_1 и x_2 — два произвольных числа из этого интервала ($a\leqslant x_1<x_2\leqslant b$) и p и q — какие угодно положительные числа, сумма которых равна единице, то*

$$pf(x_1)+qf(x_2)<f(px_1+qx_2). \qquad (7.3.4)$$

(При $p=q=1/2$ теорема 2 переходит в теорему 1.)

Доказательство. Отметим прежде всего, что если M и N — две точки, имеющие координаты (x_1, y_1) и (x_2, y_2), а Q — точка отрезка MN, делящая этот отрезок в отношении $MQ:QN=q:p$, где $p+q=1$, то координаты точки Q равны px_1+qx_2 и py_1+qy_2. Действительно, обозначим через X_1, X_2 и X; Y_1, Y_2 и Y проекции точек M, N и Q на оси координат (рис. 4); в таком случае точки X и Y будут делить отрезки X_1X_2 и Y_1Y_2 в отношении $q:p$. Отсюда получаем [2]

$$OX=OX_1+X_1X=x_1+q(x_2-x_1)=$$
$$=(1-q)x_1+qx_2=px_1+qx_2$$

и

$$OY=OY_2+Y_2Y=y_2+p(y_1-y_2)=$$
$$=(1-p)y_2+py_1=py_1+qy_2.$$

Рассмотрим теперь снова график нашей выпуклой функции $y=f(x)$ (рис. 5). Пусть $OA=x_1$, $OB=x_2$, $AM=f(x_1)$, $BN=f(x_2)$. Согласно доказанному выше координаты точки Q, делящей отрезок MN в отношении $MQ:QN=q:p$, равны px_1+qx_2 и $pf(x_1)+qf(x_2)$; таким образом, на рис. 5 $SQ=pf(x_1)+qf(x_2)$ и $SP=f(px_1+qx_2)$. Но в силу выпуклости функции $y=f(x)$ точка Q расположена н и ж е точки P, значит,

$$pf(x_1)+qf(x_2)<f(px_1+qx_2),$$

что и требовалось доказать [3].

---

[2] На рис. 4 изображен случай, когда все четыре числа x_1, x_2, y_1 и y_2 положительны; мы предоставляем читателю самостоятельно рассмотреть иные случаи.

[3] Нетрудно видеть, что координаты л ю б о й точки отрезка MN могут быть представлены в виде $(px_1+qx_2,\ py_1+qy_2)$, где

Примеры.

а) $y=\ln x$. В этом случае неравенство (4) дает

$$p\ln x_1+q\ln x_2<\ln(px_1+qx_2).$$

Отсюда следует, что

$$x_1^p x_2^q<px_1+qx_2,\quad p,q>0,\quad p+q=1.$$

б) $y=-x^2$. Имеем

$$-px_1^2-qx_2^2<-(px_1+qx_2)^2,$$

или

$$px_1^2+qx_2^2>(px_1+qx_2)^2,$$

$\sqrt{px_1^2+qx_2^2}>px_1+qx_2$, где $p,\ q>0$ и $p+q=1$.

в) $y=-\frac{1}{x}$. Здесь получаем

$$-\frac{p}{x_1}-\frac{q}{x_2}<-\frac{1}{px_1+qx_2},$$

$\frac{p}{x_1}+\frac{q}{x_2}>\frac{1}{px_1+qx_2}$, где $x_1, x_2, p,q>0$ и $p+q=1$.

Теорему 1 можно обобщить еще и по-иному.

Теорема 3. *Если $y=f(x)$ — функция, выпуклая в интервале от a до b, и $x_1,\ x_2,\ \ldots,\ x_k$ — какие-то k значений аргумента функции в этом интервале, не все равные между собой, то*

$$\frac{f(x_1)+f(x_2)+\ \ldots\ +f(x_k)}{k}<$$

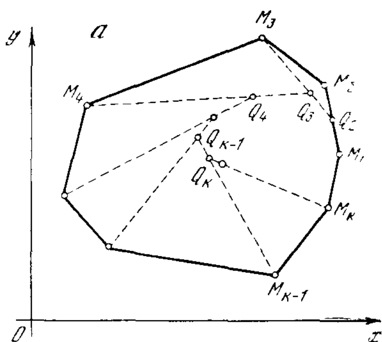
$$<f\left(\frac{x_1+x_2+\ \ldots\ +x_k}{k}\right) \qquad (7.3.5)$$

(частный случай *неравенства Иенсена* [4]).

При $k=2$ теорема 3 переходит в теорему 1.

Доказательство. Начнем с определения одного понятия, часто встречающегося в геометрических и аналитических задачах. Пусть $M_1M_2M_3\ldots$ $\ldots M_k$ — произвольный k-угольник (рис. 6, *а*); Q_2 — середина стороны M_1M_2 этого k-угольника $(M_1Q_2 : Q_2M_2 = 1/2 : 1/2)$; Q_3 — точка, делящая отрезок M_3Q_2 в отношении $2:1$ $(M_3Q_3 : Q_3Q_2 = 2/3 : 1/3)$; Q_4 — точка, делящая отрезок M_4Q_3 в отношении $3:1$ $(M_4Q_4 : Q_4Q_3 = {}^3/_4 : {}^1/_4)$; $\ldots$; наконец Q_k — точка, делящая отрезок M_kQ_{k-1} в отношении $(k-1):1$ $\left(\text{т. е. такая, что } M_kQ_k : Q_kQ_{k-1}=\frac{k-1}{k}:\frac{1}{k}\right)$.

Точка Q_k называется *центроидом* (или *центром тяжести*) k-угольника $M_1M_2\ldots M_k$. В случае треугольника



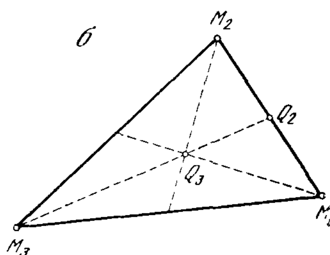


Рис. 7.3.6

$M_1M_2M_3$ (рис. 6, *б*) центроид Q_3 совпадает с точкой пересечения медиан: действительно, в этом случае Q_2 есть середина стороны M_1M_2, отрезок M_3Q_2 является медианой и точка Q_3, делящая этот отрезок в отношении $M_3Q_3 : Q_3Q_2 = 2:1$ — это точка пересечения медиан треугольника.

Докажем, что *если координаты вершин $M_1,\ M_2,\ldots,\ M_k$ k-угольника суть $(x_1,\ y_1),\ (x_2,\ y_2),\ldots,\ (x_k,\ y_k)$, то координаты центроида Q_k будут равны* $(x_1+x_2+\ldots+x_k)/k$ и $(y_1+y_2+\ldots$ $\ldots+y_k)/k$ [5].

(свои для каждой точки отрезка!) числа $p,\ q$ таковы, что $p>0,\ q>0,\ p+q=1$. Таким образом, неравенство (4) утверждает, что вся хорда MN расположена ниже кривой $y=f(x)$, т. е. оно равносильно определению выпуклости функции.

[4] Иенсен Иоганн Людвиг (1859—1925) — датский математик.

[5] Отсюда, в частности, следует, что центроид k-угольника полностью определяется этим k-угольником и не зависит от порядка перечисления его вершин (как можно было бы думать исходя из определения центроида); в случае треугольника это обстоятельство вытекает также из совпадения центроида с точкой пересечения медиан.

Действительно, в силу предложения, приведенного в начале доказательства теоремы 2, точки Q_2, Q_3, Q_4, ..., Q_k имеют следующие координаты:

$$Q_2\left(\frac{x_1+x_2}{2},\ \frac{y_1+y_2}{2}\right);$$

$$Q_3\left(\frac{2}{3}\frac{x_1+x_2}{2}+\frac{1}{3}x_3,\ \frac{2}{3}\frac{y_1+y_2}{2}+\frac{1}{3}y_3\right),$$

или $Q_3\left(\frac{x_1+x_2+x_3}{3},\ \frac{y_1+y_2+y_3}{3}\right)$;

$$Q_4\left(\frac{3}{4}\frac{x_1+x_2+x_3}{3}+\frac{1}{4}x_4,\ \frac{3}{4}\frac{y_1+y_2+y_3}{3}+\frac{1}{4}y_4\right),$$

или $Q_4\left(\frac{x_1+x_2+x_3+x_4}{4},\ \frac{y_1+y_2+y_3+y_4}{4}\right)$,

. . . .

$$Q_k\left(\frac{k-1}{k}\frac{x_1+x_2+\ldots+x_{k-1}}{k-1}+\frac{1}{k}x_k,\ \frac{k-1}{k}\frac{y_1+y_2+\ldots+y_{k-1}}{k-1}+\frac{1}{k}y_k\right),$$

или $Q_k\left(\frac{x_1+x_2+\ldots+x_{k-1}+x_k}{k},\ \frac{y_1+y_2+\ldots+y_{k-1}+y_k}{k}\right)$.

Вернемся теперь к нашей выпуклой функции $y=f(x)$. Пусть $M_1, M_2, \ldots, M_k$ — это k последовательных точек графика

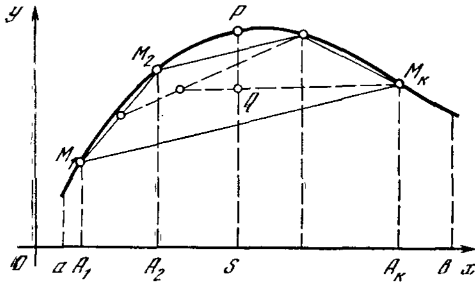


Рис. 7.3.7

этой функции, взятых в рассматриваемом интервале (рис. 7). В силу выпуклости функции k-угольник $M_1M_2\ldots M_k$ будет выпуклым и будет лежать целиком под кривой $y=f(x)$. Если абсциссы точек M_1, M_2, ..., M_k равны x_1, x_2, ..., x_k, то ординаты их, очевидно, будут равны $f(x_1)$, $f(x_2)$, ..., $f(x_k)$. Поэтому координаты центроида Q k-угольника $M_1M_2\ldots M_k$ будут равны

$$\frac{x_1+x_2+\ldots+x_k}{k}$$

и $\frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k}$

и, следовательно,

$$OS=\frac{x_1+x_2+\ldots+x_k}{k},$$

$$SQ=\frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k},$$

а

$$SP=f\left(\frac{x_1+x_2+\ldots+x_k}{k}\right)$$

(см. рис. 7). Но центроид выпуклого k-угольника лежит **внутри** k-угольника (это вытекает из самого определения центроида); следовательно, точка Q расположена **ниже** точки P и, значит,

$$\frac{f(x_1)+f(x_2)+\ldots+f(x_k)}{k}<f\left(\frac{x_1+x_2+\ldots+x_k}{k}\right),$$

что и требовалось доказать.

Это рассуждение сохраняет свою силу и в том случае, когда некоторые (но не все!) из точек M_1, M_2, ..., M_k совпадают (некоторые из чисел x_1, x_2, ..., x_k равны между собой) и k-угольник $M_1M_2\ldots M_k$ вырождается в многоугольник с меньшим числом вершин.

Примеры.

а) $y=\ln x$. Из теоремы 3 следует, что

$$\frac{\ln x_1+\ln x_2+\ldots+\ln x_k}{k}<\ln\frac{x_1+x_2+\ldots+x_k}{k},$$

или

$$\sqrt[k]{x_1x_2\ldots x_k}<\frac{x_1+x_2+\ldots+x_k}{k}$$

— *среднее геометрическое k положительных чисел, которые не все равны между собой, меньше их среднего арифметического* (**теорема о среднем геометрическом и среднем арифметическом**).

б) $y=-x^2$. В этом случае получаем

$$-\frac{x_1^2+x_2^2+\ldots+x_k^2}{k}<-\left(\frac{x_1+x_2+\ldots+x_k}{k}\right)^2,$$

или

$$\sqrt{\frac{x_1^2+x_2^2+\ldots+x_k^2}{k}}>\frac{x_1+x_2+\ldots+x_k}{k}$$

— *среднее квадратичное k произвольных положительных чисел, которые не все равны между собой, больше их среднего арифметического.*

в) $y=-\frac{1}{x}$. В этом случае теорема 3 дает

$$-\frac{1}{k}\left(\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_k}\right)< -\frac{k}{x_1+x_2+\ldots+x_k},$$

т. е.

$$\frac{1}{k}\left(\frac{1}{x_1}+\frac{1}{x_2}+\cdots+\frac{1}{x_k}\right)> \frac{k}{x_1+x_2+\ldots+x_k},$$

а, значит,

$$\frac{k}{1/x_1+1/x_2+\ldots+1/x_k}<\frac{x_1+x_2+\ldots+x_k}{k}$$

— *среднее гармоническое k положительных чисел, которые не все равны между собой, меньше их среднего арифметического.*

Наконец, докажем теорему, обобщающую теоремы 2 и 3.

Теорема 4. *Пусть $y=f(x)$ — функция, выпуклая в интервале от a до b, а $x_1, x_2, \ldots, x_k$ — какие-то k значений аргумента этой функции, не все равные между собой, взятые в рассматриваемом интервале, и $p_1, p_2, \ldots, p_k$ — k положительных чисел, сумма которых равна единице. В таком случае*

$$p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k)< f(p_1x_1+p_2x_2+\ldots+p_kx_k) \quad (7.3.6)$$

(общее неравенство Иенсена).

При $k=2$ теорема 4 переходит в теорему 2, а при $p_1=p_2=\ldots=p_k=\frac{1}{k}$ — в теорему 3.

Доказательство. Рассмотрим снова график выпуклой функции $y=f(x)$ и вписанный в этот график выпуклый k-угольник $M_1M_2\ldots M_k$, вершины которого имеют координаты $(x_1, f(x_1))$, $(x_2, f(x_2))$, $(x_3, f(x_3))$, $\ldots$, $(x_k, f(x_k))$ (рис. 8). Пусть теперь Q_2 — такая точка стороны M_1M_2 этого k-угольника, что $M_1Q_2:Q_2M_2=\frac{p_2}{p_1+p_2}:\frac{p_1}{p_1+p_2}$; Q_3 — такая точка отрезка M_3Q_2, что

$$M_3Q_3:Q_3Q_2=\frac{p_3}{p_1+p_2+p_3}:\frac{p_1+p_2}{p_1+p_2+p_3};$$

Q_4 — такая точка отрезка M_4Q_3, что

$$M_4Q_4:Q_4Q_3=\frac{p_4}{p_1+p_2+p_3+p_4}:\frac{p_1+p_2+p_3}{p_1+p_2+p_3+p_4};\ \ldots;$$

наконец, $Q_k=Q$ — такая точка отрезка M_kQ_{k-1}, что $M_kQ:QQ_{k-1}=p_k:(p_1+p_2+\ldots+p_{k-1})$ (если $p_1=p_2=\ldots$

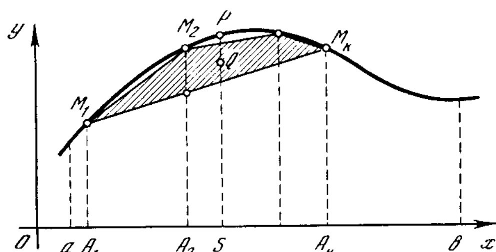


Рис. 7.3.8

$\ldots=p_k=\frac{1}{k}$, то Q — центроид k-угольника $M_1M_2\ldots M_k$). Воспользовавшись предложением, с которого мы начали доказательство теоремы 2, найдем координаты точек Q_2, Q_3, Q_4, $\ldots$, Q:

$$Q_2\left(\frac{p_1x_1+p_2x_2}{p_1+p_2},\ \frac{p_1f(x_1)+p_2f(x_2)}{p_1+p_2}\right);$$

$$Q_3\left(\frac{p_1+p_2}{p_1+p_2+p_3}\frac{p_1x_1+p_2x_2}{p_1+p_2}+\frac{p_3}{p_1+p_2+p_3}x_3,\ \frac{p_1+p_2}{p_1+p_2+p_3}\frac{p_1f(x_1)+p_2f(x_2)}{p_1+p_2}+\frac{p_3}{p_1+p_2+p_3}f(x_3)\right),$$

или $Q_3\left(\frac{1}{p_1+p_2+p_3}(p_1x_1+p_2x_2+p_3x_3),\ \frac{1}{p_1+p_2+p_3}[p_1f(x_1)+p_2f(x_2)+p_3f(x_3)]\right)$,

. . . .

$$Q\left(\frac{1}{p_1+p_2+\ldots+p_{k-1}+p_k}\times(p_1x_1+p_2x_2+\ldots+p_{k-1}x_{k-1}+p_kx_k),\ \frac{1}{p_1+p_2+\ldots+p_{k-1}+p_k}\times[p_1f(x_1)+p_2f(x_2)+\ldots+p_{k-1}f(x_{k-1})+p_kf(x_k)]\right),$$

или, иначе,

$$Q(p_1x_1+p_2x_2+\ldots+p_kx_k,\; p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k)),$$

так как $p_1+p_2+\ldots+p_k=1$. Таким образом, на рис. 8

$$SQ=p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k),$$
$$OS=p_1x_1+p_2x_2+\ldots+p_kx_k,$$
$$SP=f(p_1x_1+p_2x_2+\ldots+p_kx_k).$$

Так как точка Q расположена ниже точки P (ибо весь k-угольник $M_1M_2\ldots M_k$ лежит под кривой $y=f(x)$, а Q — внутренняя точка этого k-угольника), то

$$p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k)<f(p_1x_1+p_2x_2+\ldots+p_kx_k),$$

что и требовалось доказать [6].

Примеры.

а) $y=\ln x$. В этом случае получаем

$$p_1\ln x_1+p_2\ln x_2+\ldots+p_k\ln x_k<\ln(p_1x_1+p_2x_2+\ldots+p_kx_k),$$

откуда, потенцируя, имеем

$$x_1^{p_1}x_2^{p_2}\ldots x_k^{p_k}<p_1x_1+p_2x_2+\ldots+p_kx_k,$$

где $p_1, p_2, \ldots, p_k\geqslant 0$ и $p_1+p_2+\ldots+p_k=1$

(обобщенная теорема о среднем геометрическом и среднем арифметическом).

б) $y=-x^2$. Имеем

$$-p_1x_1^2-p_2x_2^2-\ldots-p_kx_k^2<-(p_1x_1+p_2x_2+\ldots+p_kx_k)^2,$$

или

$$\sqrt{p_1x_1^2+p_2x_2^2+\ldots+p_kx_k^2}>p_1x_1+p_2x_2+\ldots+p_kx_k,\quad p_1,p_2,\ldots,p_k\geqslant 0,$$
$$p_1+p_2+\ldots+p_k=1$$

(обобщенная теорема о среднем квадратичном и среднем арифметическом).

в) $y=-\frac{1}{x}$. Здесь теорема 4 дает

$$-\frac{p_1}{x_1}-\frac{p_2}{x_2}-\ldots-\frac{p_k}{x_k}<-\frac{1}{p_1x_1+p_2x_2+\ldots+p_kx_k},$$

откуда легко получить

$$\frac{1}{p_1/x_1+p_2/x_2+\ldots+p_k/x_k}<p_1x_1+p_2x_2+\ldots+p_kx_k,\quad p_1,p_2,\ldots,p_k\geqslant 0,$$
$$p_1+p_2+\ldots+p_k=1.$$

Упражнения

7.3.1. Докажите, что следующие функции являются выпуклыми:

а) $y=-x^n$ при $n>1$ и $x>0$; б) $y=x^m$, где $0<m<1$ и $x>0$; в) $y=-\frac{1}{x^k}$, где $k>0$ и $x>0$; г) $y=-x\log x$ при $x>0$; д) $y=-x\log x-(1-x)\log(1-x)$ при $0<x<1$ (основание системы логарифмов в упр. г) и д) — любое большее единицы число).

7.3.2. Выпишите неравенства (1), (4), (5) и (6), где $f(x)$ есть: а) функция упр. 1а); б) функция упр. 1б); в) функция упр. 1в); г) функция упр. 1г).

7.3.3. а) Докажите, что среднее геометрическое двух (положительных) чисел есть среднее геометрическое их среднего арифметического и среднего гармонического.

б) Выведите из результата упр. а) и неравенства (2) неравенство (3).

§ 4. Вычисление площадей

В гл. 3 было показано, что определенный интеграл $\int_a^b f(x)\,dx$ можно понимать как площадь криволинейной трапеции, ограниченной сверху линией $y=f(x)$, снизу — осью x, а слева и справа — вертикальными основаниями трапеции, т. е. прямыми $x=a$ и $x=b$ (рис. 1; здесь мы для простоты считаем, что $f(x)>0$ и $a<b$). Поэтому умение находить определенные интегралы дает возможность вычислять стандартными приемами различные площади, в то время как элементарная математика позволяет находить только площади прямолинейных фигур (многоугольников), а также круга и некоторых его частей (сектор, сегмент).

Найдем площадь, ограниченную графиком степенной функции $y=cx^n$ ($c>0$ и $n>0$), осью x и пря-

[6] Нетрудно видеть, что координаты любой внутренней точки k-угольника $M_1M_2\ldots M_k$ можно представить в виде $(p_1x_1+p_2x_2+\ldots+p_kx_k,\; p_1f(x_1)+p_2f(x_2)+\ldots+p_kf(x_k))$, где $p_1, p_2, \ldots, p_k>0$ и $p_1+p_2+\ldots+p_k=1$. Таким образом, неравенство (6) выражает то обстоятельство, что вписанный в график выпуклой функции многоугольник весь лежит ниже этого графика.

мой $x=x_0$ (мы считаем, что $x_0>0$; на рис. 2 $n=2$, $c=0{,}25$):

$$S=\int_0^{x_0} cx^n dx=\frac{cx^{n+1}}{n+1}\Big|_0^{x_0}=\frac{cx_0^{n+1}}{n+1}. \qquad (7.4.1)$$

Формулу (1) можно переписать так:

$$S=\frac{1}{n+1}cx_0^n x_0,$$

или, поскольку $cx_0^n=y(x_0)=y_0$,

$$S=\frac{1}{n+1}y_0x_0. \qquad (7.4.2)$$

Так как величины y и x имеют размерность длины, то из (2) следует, что S действительно имеет размерность площади (квадрата длины). Мы видим, что площадь S по порядку величины совпадает с произведением y_0x_0: от y_0x_0 величина S отличается только множителем $\frac{1}{n+1}$, при не слишком больших n по порядку величины вполне сравнимым с единицей (ср. с (5.6.6) — здесь $y_0=f_{\max}$, $x_0=b-a$, а $\frac{1}{n+1}$ — безразмерный множитель порядка 1).

В качестве следующего примера найдем площадь, ограниченную экспоненциальной кривой

$$y=ce^{-x/a}, \quad \text{где } c>0 \text{ и } a>0, \qquad (7.4.3)$$

осью x и прямыми $x=x_0$ и $x=A$ (здесь $A>x_0$; рис. 3). Эта площадь равна

$$S_A=\int_{x_0}^{A} ce^{-x/a}dx=-cae^{-x/a}\Big|_{x_0}^{A}=$$

$$=ca\left(e^{-x_0/a}-e^{-A/a}\right). \qquad (7.4.4)$$

Если A велико по сравнению с x_0, то $e^{-x_0/a}\gg e^{-A/a}$. Увеличив значение A, мы, как видно из (4), почти не изменим значение S_A. При неограниченном увеличении A величина $e^{-A/a}$ неограниченно приближается к нулю. Поэтому можно говорить о площади изображенной на рис. 3 фигуры, вовсе не ограничивая ее справа: *неограниченная* фигура, получающаяся из фигуры рис. 3 при $A\to\infty$, имеет конечную площадь

$$S_\infty=\int_{x_0}^{\infty} ce^{-x/a}dx=cae^{-x_0/a}=ay_0, \qquad (7.4.5)$$

где $y_0=y(x_0)=ce^{-x_0/a}$.

Показатель степени в (3) должен быть безразмерным. Поэтому размерность a совпадает с размерностью x,

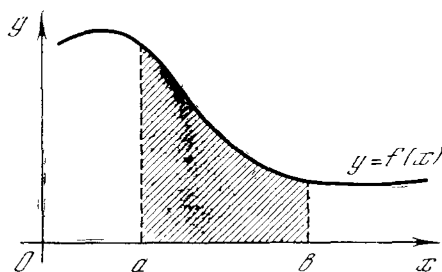


Рис. 7.4.1

т. е. a имеет размерность длины. Размерность y — также длина. Размерность S — опять-таки размерность площади.

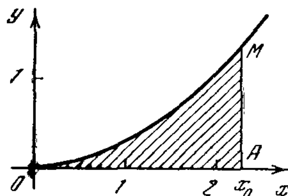


Рис. 7.4.2

Оказывается, очень просто выражается площадь под одной дугой синусоиды (рис. 4). Действительно,

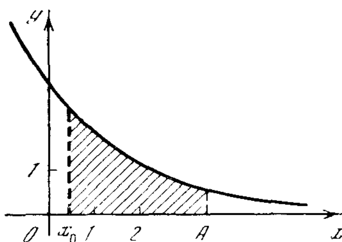


Рис. 7.4.3

эта площадь, снизу ограниченная отрезком оси x от 0 до π, равна

$$S=\int_0^{\pi}\sin x\,dx=-\cos x\Big|_0^{\pi}=2. \qquad (7.4.6)$$

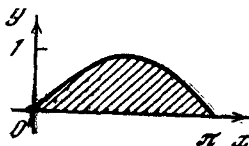


Рис. 7.4.4

Не сложнее найти и площадь под одной аркой циклоиды (рис. 5; см. также § 1.8). Если линия задана параметрическими уравнениями $x=\varphi(t)$, $y=\psi(t)$ (см. § 1.8), то формула для

площади криволинейной трапеции принимает вид

$$S=\int_a^b y\,dx=\int_\alpha^\beta y(t)\frac{dx}{dt}\,dt,$$

где $a=x(\alpha)$ и $b=x(\beta)$ — границы рассматриваемого промежутка. При $x=$

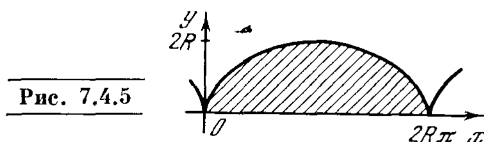


Рис. 7.4.5

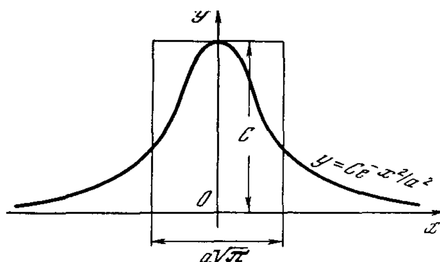


Рис. 7.4.6

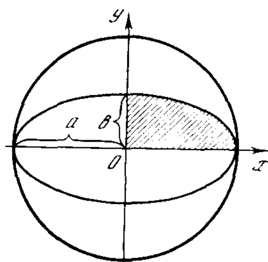


Рис. 7.4.7

$=R(t-\sin t)$, $y=R(1-\cos t)$, $\alpha=0$, $\beta=2\pi$ получаем

$$S=\int_0^{2\pi} R(1-\cos t)\,R(1-\cos t)\,dt=$$

$$=R^2\int_0^{2\pi}(1-\cos t)^2\,dt=R^2\int_0^{2\pi}(1-2\cos t+$$

$$+\cos^2 t)\,dt=R^2\int_0^{2\pi}dt-2R^2\int_0^{2\pi}\cos t\,dt+$$

$$+R^2\int_0^{2\pi}\frac{1+\cos 2t}{2}\,dt=R^2t\Big|_0^{2\pi}-2R^2\sin t\Big|_0^{2\pi}+$$

$$+\frac{R^2}{2}\left(t+\frac{1}{2}\sin 2t\right)\Big|_0^{2\pi}=R^2\cdot 2\pi-2R^2\cdot 0+$$

$$+\frac{R^2}{2}\cdot 2\pi+\frac{R^2}{4}\cdot 0=3\pi R^2 \qquad (7.4.7)$$

— площадь под одной аркой циклоиды равна утроенной площади порождающего циклоиду круга.

Результаты (2) и (5)—(7) формулируют также и по-иному. Площадь заштрихованного на рис. 2 криволинейного треугольника OAM равна $\frac{1}{n+1}OA\cdot AM=OA\left(\frac{1}{n+1}AM\right)$, т. е. совпадает с площадью прямоугольника с основанием OA и высотой $\frac{1}{n+1}AM$. В силу этого иногда говорят, что *эффективная высота* нашего криволинейного треугольника (или нашей кривой $y=cx^n$ в пределах от $x=0$ до $x=x_0$) равна $\frac{1}{n+1}$-й части его реальной высоты AM. Здесь под эффективной высотой понимается высота прямоугольника, имеющего то же основание, что и наш криволинейный треугольник, и ту же площадь. Аналогично эффективные высоты изображенных на рис. 4 и 5 криволинейных трапеций с основаниями соответственно π и $2R\pi$ равны соответственно $\frac{2}{\pi}$ ($\approx 0{,}6366$) и $1{,}5R$, в то время как реальные высоты, т. е. значения $y_{\max}$ равны соответственно 1 и $2R$. Наконец, *эффективная длина* изображенной на рис. 3 бесконечной фигуры, ограниченной экспоненциальной кривой $y=ce^{-x/a}$, осью x и прямой $x=x_0$, равна a: ведь площадь этой фигуры равна площади прямоугольника с высотой y_0 и основанием a.

Понятие эффективной длины (или эффективной высоты) кривой часто бывает полезно. Так, например, можно доказать, что изображенная на рис. 6 колоколообразная (неограниченная) кривая $y=c\exp\left(-\frac{x^2}{a^2}\right)$ ограничивает конечную площадь $S=ca\sqrt{\pi}$; другими словами, *эффективная ширина* этой кривой, высота которой равна $y_{\max}=y(0)=c$, равна $a\sqrt{\pi}\approx 1{,}77a$ (ср. [15, § 2, гл. III]).

Определим теперь площадь S эллипса с полуосями a и b (рис. 7). Конечно, эту площадь легко найти из элементарных геометрических соображений (см. текст, напечатанный ниже мелким шрифтом); мы, однако, предпочтем использовать стандартный (хоть в этом случае и более громоздкий) аппарат интегрального исчисления. Разумеется, в силу симметрии эллипса достаточно найти площадь S_1 той его части, которая лежит в первом квадранте, и резуль-

тат умножить на 4: $S=4S_1$. Для того чтобы найти S_1, выразим из уравнения $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ эллипса (см. § 1.7) y через x:
$y=\frac{b}{a}\sqrt{a^2-x^2}$

(здесь под корнем квадратным понимается арифметическое его значение, ибо в первом квадранте $y>0$). Итак,

$$S_1=\frac{b}{a}\int_0^a\sqrt{a^2-x^2}\,dx. \qquad (7.4.8)$$

Величину интеграла (7) легко найти, выполнив замену переменной по формуле $x=a\sin t$. Получим

$$\int_0^a\sqrt{a^2-x^2}dx=\int_0^{\pi/2}a\sqrt{1-\sin^2 t}\,a\cos t\,dt=$$
$$=\int_0^{\pi/2}a^2\cos^2 t\,dt=a^2\int_0^{\pi/2}\frac{1+\cos 2t}{2}\,dt=$$
$$=a^2\left[\frac{t}{2}+\frac{\sin 2t}{4}\right]_0^{\pi/2}=\frac{\pi a^2}{4}. \qquad (7.4.9)$$

Пользуясь (9), получаем из (8)
$S_1=\frac{b}{a}\frac{\pi a^2}{4}=\frac{\pi ab}{4}$.

Площадь всего эллипса
$S=\pi ab$.

Если $a=b=r$, то в полном соответствии с тем, что при $a=b=r$ эллипс переходит в окружность, получаем $S=\pi r^2$ (площадь круга).

Эллипс с полуосями a и b (где $b<a$) получается из круга радиуса a сжатием к оси x с коэффициентом сжатия $k=\frac{b}{a}$ (см. тот же рис. 7). Но нетрудно видеть, что такое сжатие переводит фигуру F площади s в фигуру F' площади $s'=ks$. В самом деле, сетку маленьких квадратиков со стороной δ и площадью δ^2, стороны которых параллельны осям координат, наше сжатие переводит в сетку прямоугольников со сторонами δ и $k\delta$ и площадью $k\delta^2$. Но сетка равных квадратиков (палетка) используется для измерения площадей: если фигура F покрывает N квадратиков сетки, то $s\approx N\delta^2$. Полученная же сжатием фигура F' покроет N прямоугольников преобразованной сетки; поэтому $s'\approx Nk\delta^2$, — откуда и следует, что $s'=ks$ (ведь δ можно выбрать сколь угодно малым, так что приближенные равенства $s\approx\approx N\delta^2$ и $s'\approx Nk\delta^2$ можно считать сколь угодно точными!). А так как площадь круга радиуса a равна πa^2, то площадь эллипса с полуосями a и b, получаемого из круга сжатием с коэффициентом $k=\frac{b}{a}$, равна $k\pi a^2=\frac{b}{a}\pi a^2=\pi ab$.

Отметим одно важное обстоятельство. Еще в гл. 3 мы отметили, что площадь (интеграл) может быть как числом положительным, так и отрицательным. Ввиду этого при нахождении площади нужна некоторая осторожность. Пусть, например, нас интересует количество краски, необходимое для того, чтобы покрасить площадку, ограниченную дугой синусоиды от $x=0$ до $x=2\pi$ и осью x (ср. рис. 3.5.2), если на окраску единицы площади требуется a граммов краски. Как указывалось выше, в этом случае нельзя сразу вычислять всю площадь как один интеграл: здесь приходится отдельно брать интегралы по участкам от 0 до π и от π до 2π.

Вообще, если подынтегральная функция $y=f(x_2)$ **меняет знак**, то для решения задачи о площади (так сказать, о «расходе краски») нужно промежуток интегрирования разбить на части, в которых $f(x)$ знака не меняет, и вычислять интеграл по отдельным частям, после чего взять сумму абсолютных величин полученных интегралов.

Упражнения

7.4.1. Найдите площадь, ограниченную осью x и дугой линии а) $y=\sin^2 x$, б) $y=\cos^2 x$, заключенной между двумя точками соприкосновения ее с осью x. [Указание. Сделайте чертеж; воспользуйтесь формулой $\sin^2 x=1/2-1/2\cos 2x$ (и $\cos^2 x=1-\sin^2 x$).]

7.4.2. Найдите площадь, ограниченную сверху линией $y=x(1-x)$, а снизу — осью x.

7.4.3. Найдите площади двух частей, на которые парабола $y=1/2\,x^2$ делит окружность $x^2+y^2=8$.

7.4.4. Найдите количество краски, необходимое для окраски площади, ограниченной: а) кривой $y=\frac{x}{1+x^2}$, осью x и вертикальными прямыми $x=1$ и $x=-1$; б) кривой $y=x^3+2x^2-x-2$ и осью x. [Указание. Предварительно постройте график функции $y=x^3+2x^2-x-2$.]

7.4.5. Чему равна площадь эллипса $\frac{x^2}{25}+\frac{y^2}{4}=1$?

§ 5*. Оценки некоторых сумм и произведений

В § 3.1 и 3.2 мы ввели (определенный) интеграл как предел некоторых сумм. Вычислив соответствующие суммы, мы можем с их помощью оценить интеграл, т. е. площадь криволинейной трапеции (см. описанные на с. 80—82 *метод прямоугольников* и *метод трапеций*). Однако связь между интегралами (площадями) и (конечными) суммами может быть использована и в обратном направлении — для оценок сумм с помощью интегралов. Эту простую идею (ср. также [15, § 2 гл. I]) мы проиллюстрируем несколькими поучительными примерами.

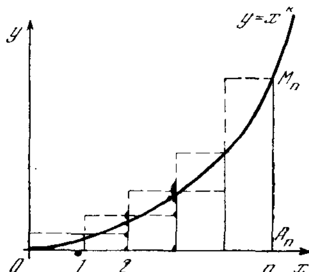


Рис. 7.5.1

Начнем с *суммы степеней натуральных чисел*:

$$S_n^{(k)}=1^k+2^k+3^k+\ldots+n^k, \qquad (7.5.1)$$

где $k>-1$. Рассмотрим функцию $y=x^k$; на рис. 1 изображен случай $k>0$; случай $0>k>-1$ мы предоставим читателю рассмотреть самостоятельно [1]. Площадь s изображенного на рис. 1 криволинейного треугольника OA_nM_n, очевидно, равна

$$s=\int_0^n x^k dx=\frac{x^{k+1}}{k+1}\Big|_0^n=\frac{n^{k+1}}{k+1}. \qquad (7.5.2)$$

С другой стороны, по правилу прямоугольников

$$1^k+2^k+\ldots+n^k>s$$
и $$0^k+1^k+\ldots+(n-1)^k<s,$$

т. е.

$$1^k+2^k+\ldots+n^k>$$
$$>\frac{n^{k+1}}{k+1}>1^k+2^k+\ldots+(n-1)^k,$$

или

$$\frac{n^{k+1}}{k+1}<S_n^{(k)}=1^k+2^k+\ldots$$
$$\ldots+n^k<\frac{n^{k+1}}{k+1}+n^k. \qquad (7.5.3)$$

Отсюда вытекает, что при больших n

$$S_n^{(k)}=1^k+2^k+\ldots+n^k\approx\frac{n^{k+1}}{k+1} \qquad (7.5.4)$$

в том смысле, что при $n\gg 1$ отношение $S_n^{(k)}\Big|\frac{n^{k+1}}{k+1}$ мало отличается от единицы: ведь

$$1\leqslant S_n^{(k)}\Big/\frac{n^{k+1}}{k+1}<1+\frac{k+1}{n} \qquad (7.5.3a)$$

((3а) получается, если разделить все части (3) на $\frac{n^{k+1}}{k+1}$).

Из (3) следует, что более точная, чем (4), оценка суммы $S_n^{(k)}$ должна иметь вид

$$S_n^{(k)}=1^k+2^k+\ldots+n^k\approx\frac{1}{k+1}n^{k+1}+cn^k,$$

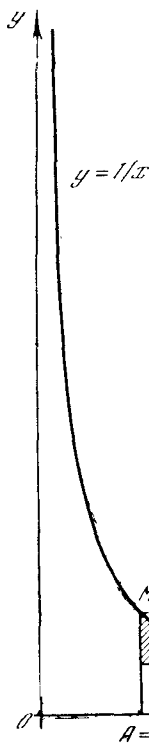


Рис. 7.5.2

---

[1] При $k<0$ нам придется рассматривать не криволинейный треугольник OA_nM_n, а криволинейную трапецию AA_nM_nM, подобную изображенной на рис. 2.

где $0 < c < 1$. И в самом деле, можно доказать, что при $k > 0$

$$S_n^{(k)} \approx \frac{1}{k+1} n^{k+1} + \frac{1}{2} n^k \qquad (7.5.4а)$$

в том смысле, что при больших n отношение $\left[S_n^{(k)} - \left(\frac{1}{k+1} n^{k+1} + \frac{1}{2} n^k\right)\right] : n^k$ будет очень мало.

По иному ведет себя при больших n сумма

$$S_n^{(-1)} = S_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}. \qquad (7.5.5)$$

Графиком функции $y = \frac{1}{x}$, к которому здесь естественно обратиться, является гипербола (рис. 2). Площадь σ криволинейной трапеции AA_nM_nM, ограниченной осью Ox, гиперболой $y = \frac{1}{x}$ и прямыми $x = 1$, $x = n$, равна

$$\sigma = \int_1^n \frac{dx}{x} = \ln x \,\Big|_1^n = \ln n, \qquad (7.5.6)$$

ибо $\ln 1 = 0$. С другой стороны, по методу прямоугольников (см. тот же рис. 2) имеем

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1} > \sigma >$$
$$> \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}, \qquad (7.5.7)$$

т. е. (ср. (5))

$$\sigma < S_n - \frac{1}{n} \quad \text{и} \quad \sigma > S_n - 1,$$

откуда

$$\sigma + \frac{1}{n} = \ln n + \frac{1}{n} < S_n <$$
$$< \ln n + 1 = \sigma + 1. \qquad (7.5.8)$$

Таким образом,

$$S_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} = \ln n + \gamma_n, \qquad (7.5.9)$$

где число γ_n заключено в пределах

$$\frac{1}{n} < \gamma_n < 1. \qquad (7.5.9а)$$

Поскольку, как легко видеть, величина γ_n представляет собой разность между площадью квадрата $AA_2M_2'M_1$ (равной 1) и суммой заштрихованных на рис. 2 площадей, то при росте n она монотонно убывает. При $n \to \infty$ число γ_n стремится к определенному пределу, который можно оценить, найдя приближенно заштрихованную на рис. 2 площадь для фиксированного (и не слишком малого) n; этот предел

$$\gamma \approx 0{,}577$$

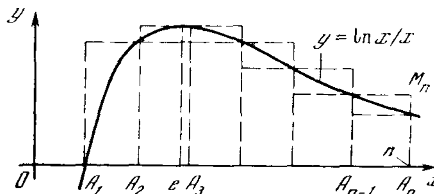


Рис. 7.5.3

называется постоянной Эйлера. Таким образом, при $n \gg 1$ имеем следующую приближенную формулу:

$$S_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} \approx$$
$$\approx \ln n + \gamma \approx \ln n + 0{,}577. \qquad (7.5.10)$$

В качестве следующего примера мы рассмотрим сумму

$$T_n = \frac{\ln 2}{2} + \frac{\ln 3}{3} + \ldots + \frac{\ln n}{n}. \qquad (7.5.11)$$

Накопленный при рассмотрении сумм (1) и (5) опыт подсказывает нам, что в этом случае уместно начать с оценки площади τ криволинейного треугольника, ограниченного линией $y = \frac{\ln x}{x}$, осью абсцисс и прямой $x = n$ (рис. 3); эту площадь τ мы сравним с суммой T_n. Площадь τ равна

$$\tau = \int_1^n \frac{\ln x}{x} dx = \int_1^n \ln x \, d(\ln x) =$$
$$= \int_0^{\ln n} u \, du = \frac{u^2}{2} \Big|_0^{\ln n} = \frac{1}{2} (\ln n)^2 \qquad (7.5.12)$$

(здесь мы ввели новую переменную $u = \ln x$).

Далее мы используем модифицированный метод прямоугольников, заключающийся в том, что площадь τ заменяется суммой площадей $n - 1$ прямоугольников с основаниями A_1A_2, ...

A_2A_3, ..., $A_{n-1}A_n$ (см. рис. 3); в качестве же высоты прямоугольника с основанием $A_{i-1}A_i$ берется в одном случае наибольшее, а в другом — наименьшее из значений функции, принимаемых ею на интервале $i-1 \leqslant x \leqslant i$; однако для этого нам понадобится подробнее изучить кривую $y = \frac{\ln x}{x}$. Очевидно, что $y(1)=0$ и $y \to 0$ при $x \to \infty$ (ср. § 6.5); найдем максимум функции $y(x)$. Поскольку равенство

$$y'(x) = \frac{(\ln x)'}{x} + \ln x\left(\frac{1}{x}\right)' = \frac{1}{x^2} - \frac{\ln x}{x^2} = \frac{1-\ln x}{x^2} = 0$$

имеет единственное решение $x=e$ (так как $\ln x = 1$ лишь при $x=e$) и $y'(x) > 0$ при $x < e$, а $y'(x) < 0$ при $x > e$, то график функции y имеет изображенный на рис. 3 вид[2]: при $x=e$ функция y имеет единственный максимум $\left(\text{здесь } y_{\max} = \frac{1}{e} \approx \frac{1}{2{,}7}\right)$, от $x=1$ до $x=e$ функция возрастает, а затем монотонно убывает. Таким образом, максимум функции $y(x)$ на интервале $2 \leqslant x \leqslant 3$ равен $\frac{1}{e}$; на всех других интервалах $i-1 \leqslant x \leqslant i$, где $i \geqslant 2$ — целое число, максимум $y(x)$ совпадает с одним из значений функции на конце интервала. Так мы получим следующие оценки площади τ:

$$\tau < \frac{\ln 2}{2} + \frac{1}{e} + \frac{\ln 3}{3} + \frac{\ln 4}{4} + \ldots \\ \ldots + \frac{\ln(n-1)}{n-1} \left(= T_n + \frac{1}{e} - \frac{\ln n}{n}\right),$$

и

$$\tau > \frac{\ln 1}{1} + \frac{\ln 2}{2} + \frac{\ln 4}{4} + \ldots \\ \ldots + \frac{\ln n}{n}\left(= T_n - \frac{\ln 3}{3}\right),$$

т. е.

$$\tau - \frac{1}{e} + \frac{\ln n}{n} < T_n < \tau + \frac{\ln 3}{3}.$$

Окончательно имеем (ср. (12)):

$$\frac{1}{2}(\ln n)^2 - \frac{1}{e} + \frac{\ln n}{n} < T_n < \\ < \frac{1}{2}(\ln n)^2 + \frac{\ln 3}{3}; \qquad (7.5.13)$$

другими словами, сумма T_n не может особенно значительно отличаться от величины $\frac{1}{2}(\ln n)^2$: при всех n имеет место равенство

$$T_n = \frac{1}{2}(\ln n)^2 + \delta_n,$$

где, во всяком случае, $-0{,}368 \approx -\frac{1}{e} < \delta_n < \frac{\ln 3}{3} \approx 0{,}367$. При этом, поскольку разность между площадью τ и ее оценкой снизу по формуле прямоугольников

$$\tau - \left(T_n - \frac{\ln 3}{3}\right) \left(= \frac{\ln 3}{3} - \delta_n\right)$$

при росте n лишь увеличивается, то выражение $\frac{\ln 3}{3} - \delta_n$ монотонно растет при $n \to \infty$, оставаясь все время меньше $\frac{\ln 3}{3} - \left(-\frac{1}{e}\right) \approx 0{,}735$. Отсюда следует, что это выражение при $n \to \infty$ стремится к определенному пределу, а, значит, при $n \to \infty$ стремится к определенному пределу (заключенному между $-0{,}37$ и $+0{,}37$) также и δ_n.

Этот метод применим и к оценке величины $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$. Число $\ln(n!) = \ln(1 \cdot 2 \cdot 3 \cdot \ldots \cdot n)$ запишем так:

$$\ln(n!) = \ln 1 + \ln 2 + \ln 3 + \ldots + \ln n. \qquad (7.5.14)$$

Для оценки стоящей в правой части (14) суммы рассмотрим криволинейный треугольник AA_nM_n, ограниченный осью Ox, линией $y = \ln x$ и прямой $x = n$ ($n > 1$ — натуральное число; рис. 4). Площадь σ_n этого треугольника, очевидно, равна

$$\sigma_n = \int_1^n \ln x\, dx = x \ln x \Big|_1^n - \int_1^n x\, d(\ln x) = \\ = x \ln x \Big|_1^n - \int_1^n x \frac{dx}{x} = x\ln x \Big|_1^n - \int_1^n dx = \\ = (x \ln x - x)\Big|_1^n = n \ln n - n + 1$$

---

[2] То, что значению $x=e$, для которого $y'(x)=0$, отвечает именно максимум функции $y(x)$, легко установить и без обращения к анализу знаков производной $y'(x)$ при разных значениях x из соображений здравого смысла (ср. со сказанным на с. 187).

(ср. § 5.4). Но, вписав в наш криволинейный треугольник $A_1A_nM_n$ (прямолинейный) треугольник $A_1A_2M_2$ и $n-2$ трапеций $A_2A_3M_3M_2$, $A_3A_4M_4M_3$, ..., $A_{n-1}A_nM_nM_{n-1}$ ($A_1=A$, $A_2, \ldots, A_n$ — точки оси x с абсциссами $1, 2, \ldots, n$, а $M_2, \ldots, M_n$ — отвечающие им точки линии $y=\ln x$), найдем, что σ_n больше суммы s_n площадей этого треугольника и трапеций, равной

$$s_n = {}^1/_2 \ln 2 + {}^1/_2(\ln 2+\ln 3) + {}^1/_2(\ln 3 + \ln 4) + \ldots + {}^1/_2[\ln(n-1)+\ln n] = \ln 2 + \ln 3 + \ldots + \ln(n-1) + {}^1/_2 \ln n = (\ln 1 + \ln 2 + \ln 3 + \ldots + \ln n) - {}^1/_2 \ln n = \ln(n!) - {}^1/_2 \ln n. \qquad \text{(7.5.15а)}$$

С другой стороны, площадь σ_n криволинейного треугольника меньше суммы S_n площадей $n-2$ трапеций со средними линиями A_iM_i, где $i=2, 3, \ldots, n-1$, и высотами длины 1 (ими служат отрезки с концами $(i-{}^1/_2,\ 0)$ и $(i+{}^1/_2,\ 0)$, где i пробегает те же значения; верхней боковой стороной трапеции служит отрезок касательной к линии $y=\ln x$ в точке M_i), дополненной малой трапецией с высотой $AA_{1/2}$ длины ${}^1/_2$ и средней линией $A_{5/4}M_{5/4}$ (здесь $A_{5/4}=({}^5/_4,\ 0)$, а $M_{5/4}$ — отвечающая $A_{5/4}$ точка линии $y=\ln x$), а также прямоугольником $A_{n-1/2}A_nM_nM'_{n-1/2}$ ($A_{n-1/2}=(n-{}^1/_2,\ 0)$, $M'_{n-1/2}=(n-{}^1/_2, \ln n)$). Ясно, что

$$S_n = \frac{1}{2}\ln\frac{5}{4} + 1\cdot\ln 2 + 1\cdot\ln 3 + \ldots + 1\cdot\ln(n-1) + \frac{1}{2}\ln n = \ln 2 + \ln 3 + \ldots + \ln(n-1) + \ln n + \frac{1}{2}\ln\frac{5}{4} - \frac{1}{2}\ln n = \ln(n!) - \frac{1}{2}\ln n + \frac{1}{2}\ln\frac{5}{4}. \qquad \text{(7.5.15б)}$$

Так как $s_n < \sigma_n < S_n$, то

$$\ln(n!) - \frac{1}{2}\ln n < n\ln n - n + 1 < \ln(n!) - \frac{1}{2}\ln n + \frac{1}{2}\ln\frac{5}{4},$$

или

$$\ln(n!) > n\ln n + \frac{1}{2}\ln n - n + 1 - \frac{1}{2}\ln\frac{5}{4}$$

и

$$\ln(n!) < n\ln n + \frac{1}{2}\ln n - n + 1,$$

т. е.

$$\left(1-\frac{1}{2}\ln\frac{5}{4}\right) + \ln\left(n^n\sqrt{n}\,e^{-n}\right) < \ln(n!) < 1 + \ln\left(n^n\sqrt{n}\,e^{-n}\right).$$

Другими словами,

$$\sqrt{\frac{4}{5}}\,en^n\sqrt{n}\,e^{-n} < n! < en^n\sqrt{n}\,e^{-n}. \qquad \text{(7.5.16)}$$

Формула (16) показывает, что при больших n величина $n!$ близка к $C\sqrt{n}\,n^ne^{-n}$,

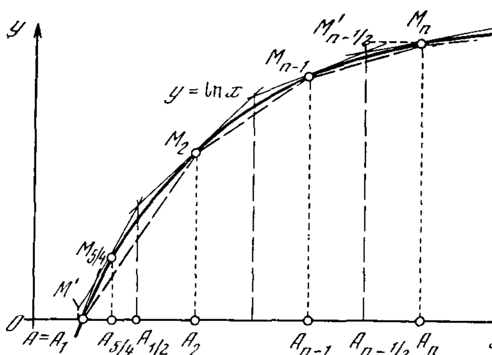


Рис. 7.5.4

где число C заключено в пределах между $e\approx 2{,}72$ и $\sqrt{\frac{4}{5}}\,e \approx 2{,}43$. При этом из того, что разность $\sigma_n - s_n$ при $n\to\infty$ монотонно возрастает, нетрудно, как и выше, вывести, что отношение $n!:(\sqrt{n}\,n^ne^{-n})$ при $n\to\infty$ стремится к определенному пределу (заключенному между $\sqrt{\frac{4}{5}}\,e$ и e, т. е. между 2,43 и 2,72). Можно доказать, что этот предел равен $\sqrt{2\pi}\approx 2{,}507$. Таким образом, мы приходим к так называемой *формуле Стирлинга*[3]

$$n! \simeq \sqrt{2\pi n}\,n^ne^{-n}, \qquad \text{(7.5.17)}$$

в которой приближенное равенство означает, что при $n\to\infty$ отношение $n!:\sqrt{2\pi n}\,n^ne^{-n}$ стремится к единице (а при большом n оно весьма близко к 1).

[3] Джеймс Стирлинг (1699—1770) — шотландский математик.

Формулу Стирлинга (17) можно обосновать еще и по-другому. Вычислим интеграл

$$I_n=\int_0^\infty x^n e^{-x}dx, \qquad (7.5.18)$$

где n — целое положительное число. Интеграл I_n можно представить себе как площадь неограниченной фигуры [4], заключенной между положительным лучом оси Ox и графиком функции $y=x^n e^{-x}$ (см. рис. 5, отвечающий случаю $n=4$). Для вычисления интеграла I_n применим метод интегрирования по частям,

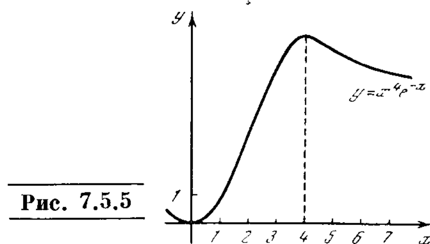


Рис. 7.5.5

полагая $e^{-x}dx=dg$, $x^n=f$; значит, $g=-e^{-x}$, $df=nx^{n-1}dx$. Таким образом, получаем

$$\int_0^\infty x^n e^{-x}dx=(-x^n e^{-x})\Big|_0^\infty+\int_0^\infty nx^{n-1}e^{-x}dx.$$

В § 6.5 было установлено, что $x^n e^{-x}=\dfrac{x^n}{e^x}\to 0$ при $x\to\infty$. Так как $x^n e^{-x}=0$ и при $x=0$, то $-x^n e^{-x}\Big|_0^\infty=0$; следовательно,

$$\int_0^\infty x^n e^{-x}dx=n\int_0^\infty x^{n-1}e^{-x}dx,$$

т. е.

$$I_n=nI_{n-1}.$$

Применив интегрирование по частям к I_{n-1}, получим точно так же $I_{n-1}=(n-1)I_{n-2}$ и т. д. Поэтому

$$I_n=n(n-1)(n-2)\cdot\ldots\cdot 3\cdot 2\cdot 1\cdot I_0,$$

где $I_0=\int_0^\infty e^{-x}dx=-e^{-x}\Big|_0^\infty=-0+1=1.$

[4] Заметьте, что при любом (сколь угодно большом!) n величина $y=x^n e^{-x}\to 0$ при $x\to\infty$ (см. § 6.5: этим мы еще воспользуемся ниже), без чего, разумеется, площадь I_n никак не могла бы существовать.

Таким образом,

$$I_n=\int_0^\infty x^n e^{-x}dx=$$
$$=n(n-1)(n-2)\cdot\ldots\cdot 3\cdot 2\cdot 1=n! \qquad (7.5.19)$$

При больших значениях n можно получить хорошее приближенное выражение для I_n, находя значение $x=x_{\max}$, при котором подынтегральная функция $x^n e^{-x}$ достигает максимума. Это значение находим из условия

$$(x^n e^{-x})'=nx^{n-1}e^{-x}+x^n(-e^{-x})=$$
$$=(n-x)x^{n-1}e^{-x}=0,$$

откуда $x_{\max}=n$. Высота этого максимума, очевидно, равна $y_{\max}=(x^n e^{-x})_{\max}=n^n e^{-n}$. Далее удается, записав $y=e^{f(x)}$, где $f(x)=n\ln x-x$, и разложив $f(x)$ в ряд Тейлора (6.1.18) вблизи значения $x=x_{\max}$ $(=n)$, оценить *эффективную ширину* нашей (неограниченной) фигуры — ширину такого прямоугольника с высотой $y_{\max}$, площадь которого равна площади I_n рассматриваемой фигуры. Эта ширина оказывается приближенно равной $\sqrt{2\pi n}$ (см. [15, гл. III, § 2 и 3]) Так мы снова приходим к формуле Стирлинга:

$$n!\simeq\sqrt{2\pi n}\,n^n e^{-n} \quad \text{при } n\gg 1.$$

§ 6*. Еще о натуральном логарифме

Остановимся на вопросе о натуральном логарифме. В гл. 3 интегрирование степенных функций привело нас к следующему результату:

$$\int x^n dx=\frac{1}{n+1}x^{n+1}+C \quad \text{при } n\neq -1; \qquad (7.6.1)$$

$$\int\frac{dx}{x}=\ln x+C, \qquad (7.6.2)$$

где, разумеется, величина C может быть любой, вовсе не обязательно одной и той же в формулах (1) и (2).

Степень $n=-1$ оказалась почему-то выпадающей из общей закономерности. Как понять этот результат? На первый взгляд, здесь содержится противоречие. Представим себе серию степенных кривых: $y=x^{-0,9}$, $y=x^{-0,99}$, $y=x^{-1}$, $y=x^{-1,01}$, $y=x^{-1,1}$.

Все эти кривые близки одна к другой: кривая с показателем степени $n=-1$ качественно ничем не выделяется по сравнению с соседними, например по сравнению с кривыми, для которых $n=-0,99$ при $n=-1,01$. Значит, и

площадь под кривой $y=x^n$ (в каких-то фиксированных пределах $a<x<b$) должна изменяться плавно, когда показатель проходит через значение $n=-1$. Но эта площадь выражается определенным интегралом, который при $n=-1$ имеет совсем другой вид, чем при $n=-0{,}99$ или $n=-1{,}01$.

Чтобы устранить это кажущееся противоречие, надо показать, что при n, близком к -1, найденный по формуле (1) определенный интеграл близок к логарифмическому выражению (2). Проделаем необходимые вычисления:

при $n=-1$ и $a, b>0$

$$S=\int_a^b \frac{dx}{x}=\ln\frac{b}{a}; \qquad (7.6.3)$$

при $n=-1+\varepsilon$

$$S=\int_a^b x^{-1+\varepsilon}dx=\frac{1}{\varepsilon}(b^\varepsilon-a^\varepsilon). \qquad (7.6.4)$$

Если показатель n близок к -1, то величина ε мала и выражение (4) для площади S приобретает вид, неудобный для вычислений. Любое число в нулевой степени равно 1; значит, если подставить в (4) значение $\varepsilon=0$, то будем иметь $b^\varepsilon-a^\varepsilon=1-1=0$. Таким образом, при $\varepsilon=0$ и числитель, и знаменатель в формуле (4) для S обращаются в нуль!

Здесь уже содержится ответ на вопрос, который часто задают учащиеся, сравнивая неопределенные интегралы (1) и (2): правая часть (1) содержит $n+1$ в знаменателе и, на первый взгляд, кажется, что при $n\to-1$ она стремится к бесконечности; логарифм же конечен. На самом деле бесконечность интеграла (1) — фиктивная: она исчезает при переходе к определенному интегралу, а, значит, от нее можно избавиться выбором соответствующей константы интегрирования C (например, $C=-\frac{1}{n+1}=-\frac{1}{\varepsilon}$).

Как упростить выражение для площади, в которое входят a^ε и b^ε? Как вычислять степени с малыми показателями? Напишем тождества

$$a=e^{\ln a},\quad a^\varepsilon=e^{\varepsilon\ln a}$$

и воспользуемся основным свойством числа e — тем, что при малых $r\ll 1$ имеем $e^r\approx 1+r$ (с точностью до величин второго и более высокого порядка малости — величин порядка малости r^2, r^3, ...). Если n близко к -1, т. е. $n=-1+\varepsilon$, где добавка ε очень мала, то можно применить написанную выше формулу. Получим

$$S=\frac{1}{\varepsilon}(b^\varepsilon-a^\varepsilon)=\frac{1}{\varepsilon}(e^{\varepsilon\ln b}-e^{\varepsilon\ln a})\approx$$
$$\approx\frac{1}{\varepsilon}[(1+\varepsilon\ln b)-(1+\varepsilon\ln a)]=$$
$$=\ln b-\ln a=\ln\frac{b}{a}. \qquad (7.6.5)$$

Вот парадокс и разрешился! В пределе для малого ε величина ε сократилась; оказалось, что степенная формула (1) для интеграла приводит к ответу, совпадающему с логарифмической формулой (2). Площадь S под кривой $y=x^n$ оказывается плавно зависящей от показателя n и тогда, когда n проходит через «особое» значение $n=-1$.

Заодно мы научились вычислять с большой точностью малые степени x^ε при $\varepsilon\ll 1$. Из школьного курса математики мы знаем, что $x^0=1$; теперь же мы узнали, насколько отличается x^ε от 1, если ε близко к нулю, но все же несколько отличается от нуля.

К рассмотрению интеграла $J=\int x^{-1}dx$ можно подойти и с других позиций.

Представим себе человека, не знающего, что такое логарифмы. Тем не менее он уже знаком с интегральным исчислением и, имея интеграл, хочет изучить его свойства. Итак, рассмотрим

$$J(a,\ b)=\int_a^b\frac{dx}{x}. \qquad (7.6.6)$$

Замечательное свойство интеграла (6) состоит в том, что *значение* $J(a,\ b)$ *не меняется, если* a *и* b *умножить на какую-то (одинаковую для* a *и* b*) величину* k. Что касается площади под линией $y=x^{-1}=\frac{1}{x}$, то можно сказать, что при замене пределов a и b интеграла (6) на ka и kb в k раз увеличится длина полоски, площадь которой мы вычисляем, но одновременно в k раз уменьшится и ее высота (здесь мы счи-

таем $k>1$; см. рис. 1); площадь полоски при этом не изменится.

Этот результат легко усмотреть и не прибегая к геометрии. Введем переменную $z=kx$, $x=\frac{z}{k}$, $dx=\frac{1}{k}dz$. Для новой переменной изменятся и пределы интегрирования: $x=a$ перейдет в $z=ka$, соответственно $x=b$ — в $z=kb$. Далее имеем

$$J(a,\ b)=\int_a^b \frac{dx}{x}=\int_{ka}^{kb} \frac{\frac{1}{k}\,dz}{\frac{1}{k}\,z}=$$

$$=\int_{ka}^{kb} \frac{dz}{z}=\int_{ka}^{kb} \frac{dx}{x}=J(ka,\ kb).$$

В конце цепочки формул мы воспользовались тем, что переменная интегрирования «немая», ее можно обозначать любой буквой.

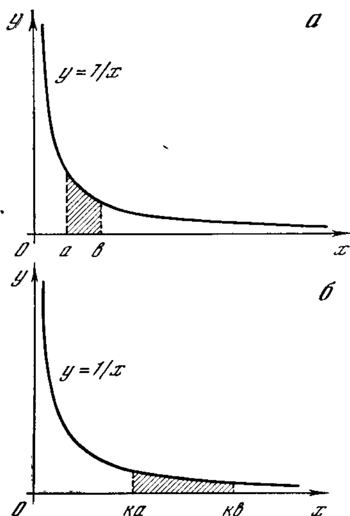


Рис. 7.6.1

Если функция двух переменных $J(a,\ b)$ не меняется при умножении обеих переменных на константу k, то эта функция зависит только от о т н о ш е н и я переменных:

$$J(a,\ b)=F(t),\quad \text{где } t=\frac{b}{a}. \qquad (7.6.7)$$

Действительно, пусть $J(1,\ t)=F(t)$; тогда, поскольку $J(a,\ b)=J(ka,\ kb)$ при любом k, положив $k=\frac{1}{a}$, получим

$$J(a,\ b)=J\left(\frac{1}{a}a,\ \frac{1}{a}b\right)=$$

$$=J\left(1,\ \frac{b}{a}\right)=J(1,\ t)=F(t).$$

Но мы знаем, что J есть интеграл. Пользуясь самыми общими свойствами интеграла, мы сумеем получить важное свойство функции $F(t)$. Положим $b=ac^m$, где m — целое число, и разобьем весь интервал интегрирования на m частей. Интеграл $J(a,\ b)$ равен сумме интегралов:

$$J(a,\ b)=J(a,\ ac)+J(ac,\ ac^2)+$$
$$+J(ac^2,\ ac^3)+\ldots+J(ac^{m-1},\ b). \qquad (7.6.8)$$

Но все интегралы в правой части (8) одинаковы! — ведь для каждого интеграла одинаковы (равны c) отношения верхнего и нижнего пределов (напомним, что $b=ac^m$). Итак,

$$J(a,\ b)=J(a,\ ac^m)=F\left(\frac{b}{a}\right)=F(c^m);$$

$$J(a,\ ac)=J(ac,\ ac^2)=\ldots=F(c),$$

и, значит,

$$F(c^m)=mF(c). \qquad (7.6.9)$$

Мы видим, что функция $F(t)$, определенная как интеграл $\int_1^t \frac{dx}{x}$ с нижним пределом интегрирования, равным 1, обладает следующим замечательным свойством: при возведении аргумента в степень m (переход $c\to c^m$) функция умножается на число m. Но ведь это есть основное свойство логарифма! Если обозначить отношение $F(d):F(c)$ (где c фиксировано, а d может меняться) через $f(d)$, то

$$f(c^m)=\frac{F(c^m)}{F(c)}=m,$$

т. е. если $d=c^m$, то $f(d)=m$; другими словами, число $f(d)$ равно той степени, в которую надо возвести постоянное фиксированное число c, для того чтобы получить число d. Но это означает, что $f(d)$ есть логарифм числа d (вспомните обычное определение логарифма!)[1]:

$$f(d)=\log_c d.$$

Нужно ли так подробно рассматривать простую формулу $\int x^{-1}dx=\ln x+C$? Для многих вычислений это не нужно — и в большинстве учебников

[1] По поводу дальнейших деталей см., например: *Маркушевич А. И.* Площади и логарифмы. М.: Наука, 1979.

и не делается. Конечно, полезно знать, как быть, если показатель степени в интеграле слева близок к -1, но не равен этой величине точно, какова погрешность при замене показателя $-1+\varepsilon$ на -1; этому посвящена первая часть настоящего параграфа. Но кроме этого, в конце концов чисто технического момента, здесь есть еще и идейная сторона, ради которой написана заключительная часть параграфа: функцию (в данном случае логарифмическую функцию) можно о п р е д е л я т ь с помощью интеграла!

В школе мы чаще всего встречаемся с функциями, заданными алгоритмом — правилом вычисления их значений. Что означает запись $f(x)=x^3$? Здесь нам задано правило: умножь x на x и еще раз на x — получишь $f(x)$. Исходя из этого правила, можно изучать свойства функции $f(x)$; так, например, ясно, что если x увеличить вдвое, то x^3 увеличится в 8 раз и т. п. Но функция может быть задана интегралом [2], и ее свойства можно изучить, даже не зная, как найти ее прямо (не путем интегрирования, а с помощью тех или иных алгебраических преобразований или действий). Так в теории вероятностей большую роль играет функция

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt. \qquad (7.6.10)$$

В теории колебаний зависимость времени от положения частицы дается функцией

$$t=\mathrm{const}\int_0^x \frac{dx}{\sqrt{x_0^2-x^2}}. \qquad (7.6.11)$$

К счастью, интеграл в правой части (11) берется в конечном виде:

$$t=k\arcsin\frac{x}{x_0},$$

и эту зависимость можно обратить, т. е. от зависимости $t=t(x)$ перейти к зависимости $x=x(t)$; при этом получаем $x=x_0\sin\omega t$ $\left(\text{здесь }\ \omega=\frac{1}{k}\right)$. Но в более сложных случаях в теории колебаний мы приходим к интегралу

$$t=\int\frac{dx}{\sqrt{ax^4+bx^3+cx^2+fx+g}}. \qquad (7.6.12)$$

Этот интеграл уже нельзя «взять», т. е. нельзя выразить через простые функции, но его можно исследовать: определить его свойства, составить таблицы.

Ради этого обобщения понятия функции и ради примера изучения свойств функции, заданной интегралом $\int\frac{dx}{x}$, читатель должен был пройти длинную цепь рассуждений, приведенных выше.

Упражнение

7.6.1. Выведите из определения (6), (7) функции $F(t)$ ее основное свойство: $F(ab)=F(a)+F(b)$ для любых положительных a и b.

[2] Заметим, что еще в начале нашего века известный немецкий математик Ф. Клейн предлагал даже в школе о п р е д е л я т ь логарифм формулой $\int_1^a\frac{dx}{x}=\ln a$.

§ 7. Средние значения

С помощью интеграла можно дать точное определение с р е д н е г о для величины, являющейся функцией какой-то переменной.

Если мы имеем величину, принимающую ряд отдельных значений, например m значений $v_1, v_2, v_3, \ldots, v_m$, то среднее значение $\bar v$ этой величины естественно определить формулой

$$\bar v=\frac{v_1+v_2+v_3+\ldots+v_m}{m}.$$

Но как определить с р е д н е е з н а ч е н и е ф у н к ц и и $v(t)$ переменной t, принимающей в с е значения в заданном промежутке от a до b $(a<t<b)$?

Представим себе, что $v(t)$ — это мгновенная скорость. Как определить среднее значение $\bar v(a, b)$, т. е. с р е д н ю ю с к о р о с т ь за время от a до b? Средняя скорость определяется как отношение пройденного пути к затраченному времени:

$$\bar v(a,\ b)=\frac{z(a,\ b)}{b-a}=\frac{\int_a^b v(t)\,dt}{b-a}. \qquad (7.7.1)$$

Это определение среднего значения функции разумно и в тех случаях, когда

функция представляет собой не скорость движения, а какую угодно другую величину. Так, например, пусть $y=y(x)$ есть изображенный на плоскости x, y график какой-нибудь функции (рис. 1).

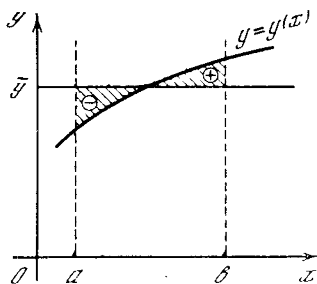


Рис. 7.7.1

Тогда $\int\limits_a^b y(x)\,dx$ есть площадь под кривой. Формула

$$\bar{y}(a,\ b)=\frac{\int\limits_a^b y(x)\,dx}{b-a},$$

$$(b-a)\,\bar{y}=\int\limits_a^b y(x)\,dx \qquad (7.7.1а)$$

значает, что $\bar{y}$ есть высота прямоугольника с основанием $b-a$, площадь

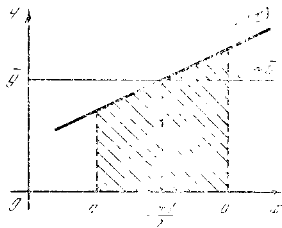


Рис. 7.7.2

которого равна площади под кривой [1]. Это значит, что на рис. 1 заштрихованная площадь над прямой $y=\bar{y}$, которая отмечена знаком плюс, в точности равна отмеченной знаком минус площади на участке, где кривая лежит ниже линии $y=\bar{y}$. График функции $y(x)$, если это не параллельная оси x прямая, обязательно проходит частью ниже, а частью выше среднего значения $\bar{y}$, определенного с помощью интеграла. Следовательно, $\bar{y}$ больше наименьшего значения $y(x)$ и меньше наибольшего значения $y(x)$ на участке усреднения $a<x<b$.

Рассмотрим примеры.

Пусть $y(x)$ есть линейная функция:

$$y=kx+m.$$

Тогда интеграл $I(a,\ b)$ представляет собой площадь трапеции (рис. 2) с высотой $b-a$, основаниями $y(a)$ и $y(b)$ и средней линией $y\left(\frac{a+b}{2}\right)$. Следовательно,

$$I(a,\ b)=\frac{y(a)+y(b)}{2}(b-a)=$$
$$=y\left(\frac{a+b}{2}\right)(b-a). \qquad (7.7.2)$$

Это выражение легко получить и без геометрических рассмотрений:

$$I(a,b)=\int\limits_a^b(kx+m)\,dx=\left(\frac{kx^2}{2}+mx\right)\Big|_a^b=$$
$$=\frac{kb^2}{2}+mb-\left(\frac{ka^2}{2}+ma\right)=$$
$$=(b-a)\left(\frac{kb}{2}+\frac{ka}{2}+m\right);$$
$$y(b)=kb+m,\ \ y(a)=ka+m,$$

и

$$y\left(\frac{a+b}{2}\right)=k\left(\frac{a+b}{2}\right)+m,$$

откуда с очевидностью следует результат (2). Таким образом, *для линейной функции*

$$\bar{y}=\frac{y(a)+y(b)}{2}=y\left(\frac{a+b}{2}\right), \qquad (7.7.2а)$$

т. е. для нее *среднее значение функции на данном участке от a до b равно среднему арифметическому значений $y(a)$ и $y(b)$ функции на краях участка.* Вот другая формулировка: *среднее значение линейной функции на некотором интервале равно значению функции в середине интервала.*

Одним из важнейших примеров линейной зависимости являются зависимости скорости от времени при равноускоренном или равнозамедленном движении, т. е. при движении тела под действием постоянной силы, скажем под действием силы тяжести, когда

---

[1] В § 4 мы называли (при $y(x)>0$ на участке $a<x<b$) величину $\bar{y}$ эффективной высотой криволинейной трапеции с площадью $S=\int\limits_a^b y\,dx$.

$v=gt+v_0$. При расчете пройденного пути используют свойства среднего линейной функции:

$$z(a,\ b)=(b-a)\left[\frac{v(b)+v(a)}{2}\right]=$$
$$=(b-a)\left(\frac{gb+ga}{2}+v_0\right).$$

Следует иметь в виду, однако, что при другой (не линейной) зависимости выражения (2а) для среднего уже не верны.

Рассмотрим далее пример квадратичной функции (параболы) $y=rx^2+px+q$. Пусть, для определенности, $r>0$; рассмотрим какую-то дугу параболы: $a<x<b$. Из рис. 3 видно [2], что

$$y\left(\frac{a+b}{2}\right)<\frac{y(a)+y(b)}{2}.$$

Обратимся теперь к интегралу $\int\limits_a^b y(x)\,dx$, т. е. к вычислению площади под параболой. Эта площадь меньше площади трапеции с основаниями A_0A и B_0B. С другой стороны, если провести через точку $C\left(\frac{a+b}{2},\ y\left(\frac{a+b}{2}\right)\right)$ касательную к параболе, то эта касательная пересечет вертикали $x=a$ и $x=b$ в точках A' и B' и образует трапецию $A_0B_0B'A'$ со средней линией $C_0C=$ $=y\left(\frac{a+b}{2}\right)$; площадь этой трапеции, очевидно, меньше площади под кривой. Таким образом, в случае параболы с $r>0$

$$(b-a)\,y\left(\frac{a+b}{2}\right)<\int\limits_a^b y(x)\,dx<$$
$$<(b-a)\,\frac{y(a)+y(b)}{2}.$$

Соответственно этому для среднего значения $\bar{y}$ в промежутке от a до b имеем

$$y\left(\frac{a+b}{2}\right)<\bar{y}<\frac{y(a)+y(b)}{2}.$$

[2] Ср. § 3. Напомним, что парабола $y=rx^2$, $r>0$, выпукла книзу, а парабола $y=rx^2+$ $+px+q$ с любыми p и q получается из параболы $y=rx^2$ параллельным переносом (см. § 1.4).

Для квадратичной функции имеет место точная формула Симпсона [3] (приводим ее без вывода, см. упр. 2), справедливая при любом знаке r:

$$\bar{y}=\frac{2}{3}\,y\left(\frac{a+b}{2}\right)+\frac{1}{3}\left[\frac{y(a)+y(b)}{2}\right]=$$
$$=\frac{1}{6}\,y(a)+\frac{2}{3}\,y\left(\frac{a+b}{2}\right)+\frac{1}{6}\,y(b). \quad (7.7.3)$$

Это выражение является хорошей приближенной формулой для расчета площади под любой плавной кривой (ср. упр. 4б).

Отметим два простых факта, относящихся к средним значениям.

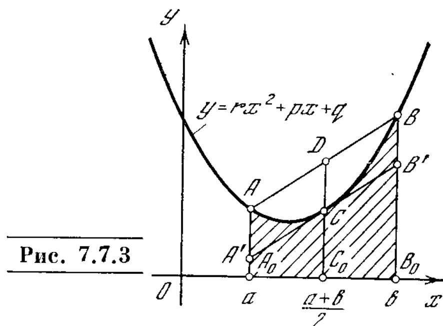


Рис. 7.7.3

1. *Среднее значение постоянной величины на любом промежутке есть сама эта постоянная.* Это ясно и из физических соображений: действительно, если мгновенная скорость не меняется, то средняя скорость за промежуток времени равна этому постоянному значению мгновенной скорости.

Совсем просто получить это и из формулы (1а):

$$\bar{C}(a,\ b)=\frac{\int\limits_a^b C\,dx}{b-a}=\frac{C(b-a)}{b-a}=C.$$

2. *Среднее значение суммы двух функций равно сумме средних значений слагаемых:*

$$\overline{y_1+y_2}=\bar{y}_1+\bar{y}_2.$$

[3] Симпсон Томас (1710—1761) — английский математик. Формула Симпсона (3) справедлива даже (см. упр. 2в) и для любой кубической функции $y=Ax^3+Bx^2+$ $+Cx+D$ (при любых A, B, C и D; если $A=0$, мы приходим к случаю квадратичной функции).

Действительно,

$$\overline{y_1+y_2}=\frac{1}{b-a}\int_a^b [y_1(x)+y_2(x)]\,dx=$$
$$=\frac{1}{b-a}\int_a^b y_1(x)\,dx+$$
$$+\frac{1}{b-a}\int_a^b y_2(x)\,dx=\bar{y}_1+\bar{y}_2.$$

Рассмотрим несколько примеров. Найдем среднее значение функции $y=\sin x$ на промежутке от $x=0$ до $x=\pi$:

$$\bar{y}(0,\ \pi)=\frac{\int_0^\pi \sin x\,dx}{\pi-0}=\frac{2}{\pi}\approx 0{,}637.$$

Среднее значение функции $y=\sin x$ на промежутке от $x=0$ до $x=b$ равно

$$\bar{y}(0,\ b)=\frac{\int_0^b \sin x\,dx}{b-0}=\frac{1-\cos b}{b}. \qquad (7.7.4)$$

А что будет, если неограниченно увеличивать число b, т. е. неограниченно увеличивать промежуток? При любом b числитель (4) не больше двух (он равен двум, если $\cos b=-1$, т. е. при $b=\pi$, 3π, 5π, 7π, . . .). Знаменатель (4) будет неограниченно увеличиваться. Поэтому можно сказать, что на *бесконечном* промежутке (к которому мы приходим, неограниченно увеличивая b) среднее значение синуса будет равно нулю (точно это означает, что чем больше промежуток, тем ближе к нулю значение среднего). Совершенно аналогично можно показать, что среднее значение функции $y=\cos x$ на бесконечном промежутке также равно нулю: ведь

$$\bar{y}(0,\ b)=\frac{\int_0^b \cos x\,dx}{b-0}=\frac{\sin x\,|_0^b}{b}=\frac{\sin b}{b}, \qquad (7.7.5)$$

и если неограниченно увеличивать b, то знаменатель (5) неограниченно увеличивается, а числитель остается не больше единицы. Следовательно, вся дробь стремится к нулю: $\bar{y}(0,\ \infty)=0$. Буквально так же получим, что и среднее значение функции $y=\cos kx$ на бесконечном промежутке тоже равно нулю.

Найдем среднее значение функции $y=\sin^2 x$ на бесконечном промежутке от $x=0$ до $x=\infty$. По известной формуле тригонометрии $\sin^2 x=\frac{1-\cos 2x}{2}$. Отсюда

$$\overline{\sin^2 x}=\overline{\tfrac{1}{2}}-\overline{\tfrac{1}{2}\cos 2x}=\frac{1}{2}-0=\frac{1}{2}.$$

Воспользовавшись формулой $\sin^2 x+\cos^2 x=1$, получаем среднее значение $\cos^2 x$ на том же промежутке:

$$\overline{\cos^2 x}=\overline{1}-\overline{\sin^2 x}=1-{}^1/_2={}^1/_2.$$

Пользоваться средними практически удобно, часто даже удобнее, чем пользоваться интегралами. По существу, эти величины равноценны: зная интеграл $I=\int_a^b y\,dx$, легко находим среднее как $\bar{y}=\frac{I}{b-a}$, а вычислив среднее $\bar{y}$, также просто находим и интеграл $I=(b-a)\,\bar{y}$.

Удобство среднего заключается в том, что $\bar{y}$ — это величина той же размерности, что и y, и, очевидно, того же порядка величины, что и значения y на исследуемом участке (ср. со сказанным по поводу формулы (5.6.6) на с. 149). Поэтому труднее пропустить ошибку в 10 раз в значении $\bar{y}$, чем такую же ошибку в значении интеграла.

Обычно предполагается, что изучающие высшую математику в совершенстве знают арифметику и алгебру и никогда не ошибаются в 10 раз или в знаке. Однако опыт показывает, что это не так! Поэтому и расчеты всегда надо вести так, чтобы уменьшить вероятность незамеченной ошибки.

Укажем еще, что определение (1) среднего $\bar{v}$ функции $v(t)$, пожалуй, ближе не к понятию среднего $\bar{v}=\frac{1}{m}(v_1+v_2+\ldots+v_m)$ ряда чисел, понимаемого как их среднее арифметическое, а к понятию «взвешенного» среднего. Пусть нам требуется определить среднюю массу $\bar{p}$ камней в куче (или среднюю цену $\bar{z}$ книг в магазине). Найдем сперва общую массу камней (общую цену книг), для чего составим сумму $P=p_1n_1+p_2n_2+\ldots+p_kn_k$ (соответственно $Z=z_1n_1+z_2n_2+\ldots+z_kn_k$), где n_i есть число «однородных» камней (книг), имеющих одну и ту же массу p_i

(одну и ту же цену z_i); после этого полученную сумму P (или Z) мы разделим на общее число N камней (книг):

$$\bar{p}=\frac{p_1n_1+p_2n_2+\ldots+p_kn_k}{N}$$

$$\left(\text{или } \bar{z}=\frac{z_1n_1+z_2n_2+\ldots+z_kn_k}{N}\right).$$

Точно так же приходится поступать во всех случаях, когда ищется среднее величин $y_1, y_2, \ldots, y_N$ и все эти величины разбиваются на k групп однородных объектов численностей $n_1, n_2, \ldots, n_k$ (разумеется, $n_1+n_2+\ldots+n_k=N$).

Вернемся теперь снова к примеру, где (непрерывно меняющаяся величина) v — это скорость; предположим, что нам дана зависимость $v=v(x)$, указывающая (мгновенную!) скорость v в каждой точке x пути. Попробуем отыскать общее время T движения со скоростью $v(x)$, требующееся, чтобы пройти от пункта $x=a$ до пункта $x=b$. Время Δt прохождения малого отрезка Δx пути будет, очевидно, равно частному $\frac{\Delta x}{v}$, где v — скорость на пути Δx, которую можно считать постоянной, ибо на этом малом пути она не успеет значительно измениться. (Все же более точной здесь будет запись $\Delta t\approx\frac{\Delta x}{v}$, поскольку даже на протяжении малого пути Δx скорость не остается одинаковой.) Общее время пути приближенно выразится суммой

$$T\approx\frac{\Delta x_1}{v_1}+\frac{\Delta x_2}{v_2}+\ldots+\frac{\Delta x_k}{v_k}, \qquad (7.7.6)$$

где $\Delta x_i=x_i-x_{i-1}$ (здесь $i=1, 2, \ldots, k$; $x_0=a, x_1, x_2, \ldots, x_k=b$ — точки, разбивающие весь путь от a до b на k малых интервалов), а за v_i можно принять как $v(x_i)$, так и $v(x_{i-1})$ (ср. с § 3.1). Точное же значение времени T выразится интегралом

$$T=\int_a^b\frac{dx}{v(x)}, \qquad (7.7.6\text{a})$$

являющимся пределом «взвешенного среднего» (6), в котором разные участки Δx_i пути берутся с разными весами $\frac{1}{v_i}$; при этом среднее значение $\frac{1}{v}$ равно частному от деления интеграла (6а) на полный путь $b-a$.

С фактом использования в формулах интегрального исчисления именно взвешенных средних связан ряд обстоятельств, которые нельзя упускать из виду. Пусть мы имеем некоторую химическую реакцию, скорость $F=F(\theta)$ которой, т. е. количество вступающего в реакцию вещества, отнесенное к единице времени, зависит от температуры θ, при которой идет реакция; сама же температура θ меняется с течением времени t, т. е. $\theta=\theta(t)$. При этом средняя температура реакции, продолжающейся от момента

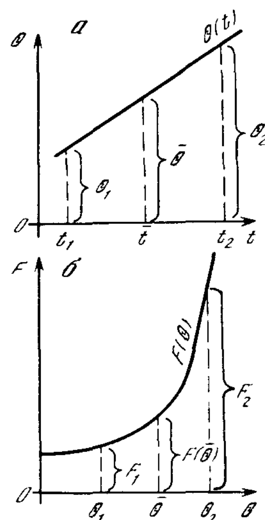


Рис. 7.7.4

$t=t_1$ до момента $t=t_2$, будет, очевидно, равна

$$\bar{\theta}=\frac{1}{t_2-t_1}\int_{t_1}^{t_2}\theta(t)\,dt; \qquad (7.7.7)$$

средняя же скорость реакции выражается формулой

$$\bar{F}=\frac{1}{t_2-t_1}\int_{t_1}^{t_2}F(\theta)\,dt. \qquad (7.7.7\text{a})$$

Не надо только думать, что средняя скорость реакции $\bar{F}$ совпадает со скоростью $F(\bar{\theta})$, характеризующей среднюю температуру $\bar{\theta}$! Так, пусть зависимость $\theta=\theta(t)$ линейная (рис. 4, *а*); тогда, очевидно, $\bar{\theta}=\frac{\theta_1+\theta_2}{2}=\theta\left(\frac{t_1+t_2}{2}\right)$. Что касается скорости F реакции, то мы предположим, что она экспоненциально растет с температурой θ, т. е. $F(\theta)=e^{k\theta}$. В этом случае в силу высокой скорости роста F (рис. 4, *б*; заметим, что

вследствие равномерности увеличения температуры θ с течением времени t зависимость $F(t)$ также имеет экспоненциальный характер) значение интеграла (7а) вообще почти целиком определяется (большими) значениями F у правого конца интервала изменения переменной θ (или t) и мало зависит от других значений F; поэтому равенство $\bar{F}=F(\bar{\theta})=\frac{1}{2}[F(\theta_1)+F(\theta_2)]$ никак не возможно. Последнее обстоятельство

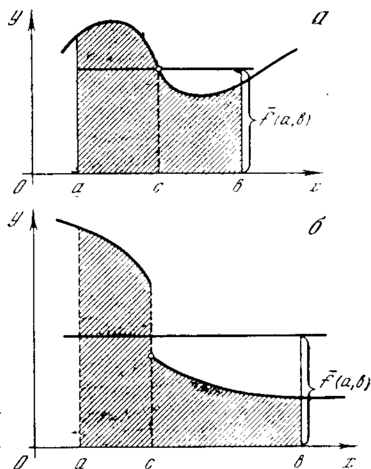


Рис. 7.7.5

целиком объясняется неравномерностью «весов» $F(\theta)$, приписываемых отдельным интервалам $\Delta\theta$ изменения θ.

В связи с вопросом о средних значениях отметим две теоремы, относящиеся к н е п р е р ы в н ы м и г л а д к и м функциям.

1. Т е о р е м а Л а г р а н ж а. *Среднее значение функции равно значению функции в какой-то точке внутри рассматриваемого интервала.* В обозначениях формулы (1а)

$$\bar{f}(a, b)=f(c), \quad a<c<b. \tag{7.7.8}$$

Нарисуем площадь, равную интегралу $\int_a^b f(x)\,dx$, и равновеликий этой криволинейной трапеции (равный ей по площади) прямоугольник, основанием которого является отрезок от a до b. Высота прямоугольника — эффективная высота трапеции — равна среднему $\bar{f}(a, b)$ функции. Очевидно, что верхнее основание прямоугольника обязательно пересекает дугу кривой $y=f(x)$ в какой-то точке $(c, f(c))$, промежуточной между концами этой дуги (рис. 5, *а*). Разумеется, здесь мы считаем функцию $y=f(x)$ н е п р е р ы в н о й; если график функции $y=f(x)$ претерпевает разрыв, то наше утверждение вполне может не выполняться (рис. 5, *б*).

Без ссылки на чертеж здесь можно рассуждать так: если M и m — н а и б о л ь ш е е и н а и м е н ь ш е е значения функции на отрезке $a \leqslant x \leqslant b$, т. е. если $M \geqslant f(x) \geqslant m$, то

$$M(b-a)=\int_a^b M\,dx \geqslant \int_a^b f(x)\,dx \geqslant$$

$$\geqslant \int_a^b m\,dx = m(b-a)$$

и, значит,

$$M \geqslant \frac{1}{b-a}\int_a^b f(x)\,dx=\overline{f(x)} \geqslant m.$$

Но отсюда уже вытекает равенство (8), ибо непрерывная функция принимает на отрезке от a до b (в некоторой точке c этого отрезка) любое значение A, промежуточное между ее наибольшим и наименьшим значениями (т. е. такое, что $M \geqslant A \geqslant m$), в частности значение

$$\frac{1}{b-a}\int_a^b f(x)\,dx.$$

Совершенно аналогично устанавливается следующее простое обобщение доказанной теоремы: *если функция $f(x)$ на отрезке $a \leqslant x \leqslant b$ непрерывна, а функция $g(x)$ сохраняет знак* (т. е. либо нигде не отрицательна на этом отрезке, либо нигде не положительна), *то*

$$\int_a^b f(x)\,g(x)\,dx=f(c)\int_a^b g(x)\,dx, \tag{7.7.9}$$

где по-прежнему $a<c<b$.

В самом деле, пусть для определенности $g(x) \geqslant 0$ при $a \leqslant x \leqslant b$. Тогда, очевидно,

$$M\int_a^b g(x)\,dx=\int_a^b Mg(x)\,dx \geqslant \int_a^b f(x)\,g(x)\,dx \geqslant$$

$$\geqslant \int_a^b mg(x)\,dx = m\int_a^b g(x)\,dx, \tag{7.7.10}$$

где снова M и m — наибольшее и наименьшее значения функции $f(x)$ на отрезке $a \leqslant x \leqslant b$.

Разделив обе части (10) на $\int\limits_a^b g(x)\,dx > 0$, получим [4]

$$M \geqslant \frac{\int\limits_a^b f(x)\,g(x)\,dx}{\int\limits_a^b g(x)\,dx} \geqslant m. \qquad (7.7.11)$$

Но непрерывная функция $f(x)$ принимает на отрезке $a \leqslant x \leqslant b$ все значения, промежуточные между самым большим и самым малым ее значениями; поэтому каждое заключенное между M и m число, — а значит, и стоящее в средней части (11) выражение — равно значению $f(c)$ функции в какой-то (промежуточной между a и b) точке c:

$$\frac{\int\limits_a^b f(x)\,g(x)\,dx}{\int\limits_a^b g(x)\,dx} = f(c). \qquad (7.7.12)$$

(Ясно, что (12) равносильно (9).)

Пример. Если $f(x)$ — произвольная непрерывная функция и a и b одного знака (т. е. оба неотрицательны или оба неположительны), то

$$\int\limits_a^b x^n f(x)\,dx = f(c)\int\limits_a^b x^n dx = \frac{b^{n+1} - a^{n+1}}{n+1} f(c), \qquad (7.7.13)$$

где число c лежит между a и b.

2. *Возьмем функцию* $y(x)$ *такую, что* $y(a) = y(b)$. Предполагается, что функция гладкая; из нашего условия, разумеется, следует, что эта функция не монотонна в рассматриваемом интервале $a < x < b$. Но если функция не монотонна, то где-то внутри участка $[a, b]$ она имеет минимум или максимум. *Значит, внутри участка при некотором $x = c$, где $a < c < b$, производная обращается в нуль*: $y'(c) = 0$. Это утверждение называется теоремой Ролля [5]. Геометрический смысл теоремы Ролля ясен из рис. 6, где изображены два варианта кривых с максимумом и минимумом (вторая кривая имеет между a и b и максимум, и минимум).

Теоремы Лагранжа и Ролля связаны между собой: теорему Ролля можно получить как следствие теоремы Лагранжа. В самом деле, запишем следующее тождество, вытекающее

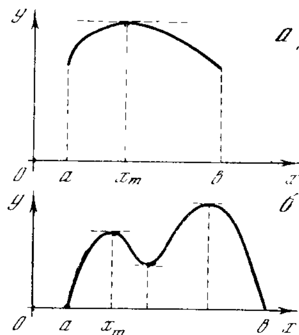


Рис. 7.7.6

из связи между производной и интегралом (см. § 3.3):

$$y(b) = y(a) + \int\limits_a^b y'(x)\,dx.$$

Отсюда следует, что

$$\overline{y'}(a, b) = \frac{\int\limits_a^b y'(x)\,dx}{b-a} = \frac{y(b) - y(a)}{b-a}.$$

Но в условиях теоремы Ролля $y(b) = y(a)$; поэтому здесь

$$\overline{y'} = 0.$$

Теперь остается только применить теорему Лагранжа к функции $y'(x)$: если среднее значение функции на интервале от a до b равно нулю, то в этом интервале есть точка c, в которой y' равно нулю.

Упражнения

7.7.1. Найдите среднее значение функции $y = x^2$ на участке от 0 до 2. Сравните это среднее значение со средним арифметическим значений функции на краях и со значением в середине промежутка.

7.7.2. Проверьте формулу Симпсона (3): а) для функции упр. 1; б) для общей квадратичной функции $y = rx^2 + px + q$; в) для кубической функции $y = Ax^3 + Bx^2 + Cx + D$.

7.7.3. Сила тяжести убывает с расстоянием r от центра Земли по закону $F = \frac{A}{r^2}$.

---

[4] Ясно, что если знакопостоянная функция $g(x)$ (т. е. функция, сохраняющая на отрезке $a \leqslant x \leqslant b$ постоянный знак) не является «патологической» (вроде, скажем, описанной на с. 425 функции $f(x)$), то равенство $\int\limits_a^b g(x)\,dx = 0$ при $b \neq a$ означает, что $g(x) \equiv 0$, — а в этом случае соотношение (9) является очевидным.

[5] Ролль Мишель (1652—1719) — французский математик.

Найдите среднее значение силы тяжести на участке от поверхности Земли (радиус R) до точки, удаленной на расстояние R от Земли (т. е. на расстояние $2R$ от центра Земли). (В этом упражнении для нахождения среднего используется «интеграл энергии» $\int F(x)\,dx$ (F — сила, а x — путь), имеющий важный физический смысл.)

7.7.4. Сравните найденное в упр. 3 значение среднего: а) со средним арифметическим значений силы на краях участка ее изменения (совпадающим со средним для линейно изменяющейся величины); б) со средним, найденным по формуле Симпсона (3) (совпадающей с истинным значением среднего для квадратичного закона изменения рассматриваемой величины).

7.7.5. Найдите среднее значение функции $y=x^n$ на промежутке от $x=0$ до $x=x_0$.

7.7.6. Найдите среднее значение функции $y=ce^{kx}$ на промежутке, в котором y меняется от $y=n$ до $y=m$; выразите это среднее значение через n и m, исключая c и k из ответа. Исследуйте полученное выражение при m, близком к n: $m=n+\nu$, $\nu \ll n$.

7.7.7. Найдите средние значения функций $y=\sin^2 x$, $y=\cos^2 x$ на промежутке: а) от $x=0$ до $x=\pi$; б) от $x=0$ до $x=\pi/4$.

7.7.8. Определите период функции $y=\sin(\omega t+\alpha)$, где ω, α — постоянные числа. Найдите среднее значение функции y^2 за ее период.

§ 8. Длина кривой

Поставим задачу: найти длину дуги s кривой $y=f(x)$ от точки, где $x=a$, до точки, где $x=b$ (рис. 1). Длину Δs малой дуги MN линии $y=f(x)$ заменяем длиной отрезка MN, соединяющего точки M и N [1]. (Мы рассматриваем только кривые без разрывов и изломов.) По теореме Пифагора

$$\Delta s \approx \sqrt{(\Delta x)^2+(\Delta y)^2}=\Delta x\sqrt{1+\left(\frac{\Delta y}{\Delta x}\right)^2}.$$

Отсюда

$$\frac{\Delta s}{\Delta x}\approx\sqrt{1+\left(\frac{\Delta y}{\Delta x}\right)^2}. \qquad (7.8.1)$$

Перейдем в (1) к пределу при $\Delta x\to 0$; отношение $\frac{\Delta y}{\Delta x}$ превратится при этом

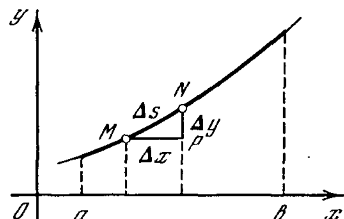


Рис. 7.8.1

в производную $y'=f'(x)$. Таким образом, получаем

$$ds=\sqrt{1+(f'(x))^2}\,dx.$$

Вся длина дуги равна

$$s=s(a,\ b)=\int_a^b\sqrt{1+(f'(x))^2}\,dx. \qquad (7.8.2)$$

Если линия задана параметрическими уравнениями $x=x(t)$, $y=y(t)$ (см. § 1.8), то из (приближенного) равенства $\Delta s\approx\sqrt{\Delta x^2+\Delta y^2}$, деля обе части его на Δt, переходя к пределу при $\Delta t\to 0$ и учитывая, что $\lim\limits_{\Delta t\to 0}\frac{\Delta x}{\Delta t}=\frac{dx}{dt}$, $\lim\limits_{\Delta t\to 0}\frac{\Delta y}{\Delta t}=\frac{dy}{dt}$, получаем $\frac{ds}{dt}=\sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2}$, или $ds=\sqrt{(x')^2+(y')^2}\,dt$; Отсюда, интегрируя от $t=\alpha$ до $t=\beta$, получаем

$$s(\alpha,\ \beta)=\int_\alpha^\beta\sqrt{(x')^2+(y')^2}\,dt, \qquad (7.8.2а)$$

где $s(\alpha,\beta)$ — длина дуги кривой, ограниченной значениями α и β параметра t.

Из-за наличия корня под знаком интеграла в (2) (и в (2а)) этот интеграл редко удается «взять», т. е. вычислить до конца, выразив величину $s=s(a,\ b)$ (или $s=s(\alpha,\ \beta)$) в виде явно выписан-

[1] Отличие длины дуги от длины соответствующей ей хорды (отрезка прямой) имеет порядок величины $|\Delta x|^3$ — и при переходе к пределу (к дифференциалам) им можно законно пренебречь. Так, например, дуга окружности радиуса r, стягивающая малый угол Δt, равна $r\Delta t$ (мера углов — радианная!), а длина соответствующей ей хорды равна $2r\sin\frac{\Delta t}{2}=2r\left(\frac{\Delta t}{2}-\frac{1}{6}\left(\frac{\Delta t}{2}\right)^3+\ldots\right)=r\Delta t-\frac{1}{24}r(\Delta t)^3+\ldots$ (см. § 6.2), так что разность длин дуги и хорды примерно равна $\frac{1}{24}r(\Delta t)^3$, т. е. имеет порядок величины $(\Delta s)^3$, где Δs — величина порядка длины дуги (или длины хорды).

ной функции пределов интегрирования a и b (или α и β).

Заметим еще, что из того же прямоугольного треугольника[2] со сторонами Δx, Δy и MN ($\approx \Delta s$), с помощью которого мы выразили Δs через Δx и Δy, имеем

$$\Delta x \approx \Delta s \cos \psi,$$

$$\Delta y \approx \Delta s \sin \psi,$$

где ψ — угол, образованный прямой MN с осью Ox. При $\Delta x \to 0$, очевидно, $\psi \to \varphi$, где φ — угол, образованный с осью Ox касательной к линии $y = f(x)$. Таким образом,

$$dx = ds \cos \varphi, \quad dy = ds \sin \varphi$$

$$\text{и} \quad ds = \sqrt{dx^2 + dy^2}, \tag{7.8.2б}$$

где употребление дифференциалов указывает, что в этих формулах $dx (= \Delta x)$, как и dy, ds, малы.

Приведем несколько примеров, когда выкладки нетрудно довести до конца.

1. Длина окружности. Будем искать длину окружности $x^2 + y^2 = R^2$ или, точнее, длину s четверти окружности, лежащей в первом квадранте; полученный результат мы затем умножим на 4.

Из уравнения окружности имеем:

$$y = \sqrt{R^2 - x^2}, \quad y' = -\frac{x}{\sqrt{R^2 - x^2}}.$$

По формуле (2)

$$s = \int_0^R \sqrt{1 + \frac{x^2}{R^2 - x^2}}\, dx = \int_0^R \frac{R\,dx}{\sqrt{R^2 - x^2}}. \tag{7.8.3}$$

Введем новую переменную t по формуле $x = R \sin t$; тогда $dx = R \cos t\, dt$ и из (3) следует

$$s = \int_0^{\pi/2} R\,dt = \frac{\pi R}{2}$$

(ср. § 5.6), откуда получаем хорошо известную формулу для длины окружности:

$$C = \frac{\pi R}{2} \cdot 4 = 2\pi R.$$

[2] Этот треугольник играл большую роль в рассуждениях Лейбница, который исходил из него при введении дифференциалов.

2. Цепная линия — это кривая, уравнение которой имеет вид

$$y = \frac{a}{2}\left(e^{x/a} + e^{-x/a}\right), \tag{7.8.4}$$

где a — постоянное число[3]. Название «цепная линия» происходит от того, что такую форму принимает гибкая и нерастяжимая тяжелая нить постоянной плотности (например, цепь), подвешенная за оба конца. График цепной линии изображен на рис. 2 (для $a = 2$).

Найдем длину дуги цепной линии от точки $x = 0$ до точки $x = x_0$. Из (4) следует, что $y' = \frac{e^{x/a} - e^{-x/a}}{2}$, поэтому

$$\sqrt{1 + (y')^2} = \sqrt{1 + \frac{e^{2x/a} - 2 + e^{-2x/a}}{4}} = \sqrt{\frac{\left(e^{x/a} + e^{-x/a}\right)^2}{4}} = \frac{e^{x/a} + e^{-x/a}}{2}$$

и, значит,

$$s = \int_0^{x_0} \frac{e^{x/a} + e^{-x/a}}{2}\, dx = \frac{a}{2}\left(e^{x/a} - e^{-x/a}\right)\Big|_0^{x_0} = \frac{a}{2}\left(e^{x_0/a} - e^{-x_0/a}\right).$$

3. Циклоида. Мы знаем, что (параметрическое) уравнение циклоиды имеет вид:

$$x = a(t - \sin t), \quad y = a(1 - \cos t),$$

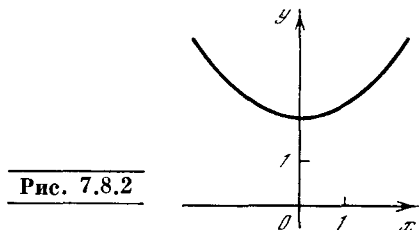


Рис. 7.8.2

где a — радиус образующего циклоиду круга (см. § 1.8). Поэтому в силу (2а)

$$ds = \sqrt{a^2(1 - \cos t)^2 + (a \sin t)^2}\, dt = a\sqrt{1 - 2\cos t + \cos^2 t + \sin^2 t}\, dt = a\sqrt{2 - 2\cos t}\, dt = a\sqrt{4 \sin^2 \frac{t}{2}}\, dt = 2a \sin \frac{t}{2}\, dt,$$

[3] Уравнение (4) можно также записать в виде $y = a \operatorname{ch} \frac{x}{a}$, где символом $\operatorname{ch} t$ обозначен так называемый *гиперболический косинус* величины t (см. § 14.4, в частности рис. 14.4.1).

15 Заказ № 1065

и дуга циклоиды, заключенная между точками, отвечающими значениям α и β параметра t, равна

$$s=2a\int_{\alpha}^{\beta}\sin\frac{t}{2}\,dt=4a\left(-\cos\frac{t}{2}\right)\Big|_{\alpha}^{\beta}=$$

$$=4a\left(\cos\frac{\alpha}{2}-\cos\frac{\beta}{2}\right).$$

Так как двум последовательным «остриям» циклоиды (см. рис. 1.8.3)

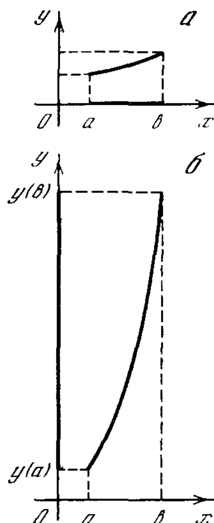


Рис. 7.8.3

отвечают значения 0 и 2π параметра t, то длина дуги одной арки циклоиды равна $8a$, т. е. она в четыре раза больше диаметра порождающего циклоиду круга.

Мы уже отмечали, что чаще всего функцию $\sqrt{1+(y'(x))^2}$ проинтегрировать в элементарных функциях трудно (или даже невозможно) из-за наличия корня. Поэтому большой интерес представляют приближенные формулы для вычисления длины дуги.

Предположим, что величина $(y'(x))^2$ мала по сравнению с единицей: $|y'(x)|\ll 1$. Тогда, пренебрегая в (2) величиной $(y'(x))^2$, получим

$$s\approx\int_a^b\sqrt{1}\,dx=b-a. \tag{7.8.5}$$

Разность $b-a$ есть длина горизонтального отрезка, концы которого суть $x=a$ и $x=b$. Формула (5) показывает, что если y' мала по абсолютной величине (кривая мало отклоняется от горизонтального отрезка), то и длина дуги этой кривой близка к длине горизонтального отрезка (рис. 3, *а*).

Если, напротив, $(y'(x))^2\gg 1$, то в (2) можно пренебречь единицей по сравнению с $(y'(x))^2$. Таким образом, здесь мы получаем

$$s\approx\int_a^b\sqrt{(y'(x))^2}\,dx=\int_a^b y'(x)\,dx=y(b)-y(a). \tag{7.8.6}$$

Формула (6) показывает, что в этом случае длина дуги кривой близка к длине вертикального отрезка, концы которого суть $y(b)$ и $y(a)$ (рис. 3, *б*). Действительно, если производная y' велика, то кривая круто поднимается вверх, а поэтому похожа на вертикальную прямую (для вертикальной прямой производная бесконечна).

Формулы (5) и (6) дают простые приближенные выражения для длины дуги, но это очень грубые приближения, которые можно найти и без обращения к формуле (2). Нетрудно получить и более точные формулы.

Пусть $|y'(x)|<1$. Сохраняя два первых члена в разложении выражения $\sqrt{1+(y'(x))^2}$ по формуле бинома Ньютона (см. § 6.4) и отбрасывая последующие члены, запишем

$$\sqrt{1+(y'(x))^2}=$$

$$=[1+(y'(x))^2]^{1/2}\approx 1+\frac{1}{2}(y'(x))^2.$$

Теперь формула (2) дает

$$s\approx\int_a^b\left[1+\frac{1}{2}(y'(x))^2\right]dx=$$

$$=(b-a)+\frac{1}{2}\int_a^b(y'(x))^2\,dx.$$

Если же $|y'(x)|>1$, то

$$\sqrt{1+(y'(x))^2}=y'(x)\sqrt{1+\frac{1}{(y'(x))^2}}.$$

К последнему выражению также можно применить формулу бинома Ньютона, поскольку $\frac{1}{(y'(x))^2}<1$. Сохраняя и в этой формуле лишь первые два члена, получим

$$y'(x)\sqrt{1+\frac{1}{(y'(x))^2}}\approx$$

$$\approx y'(x)\left[1+\frac{1}{2(y'(x))^2}\right]=y'(x)+\frac{1}{2y'(x)}. \tag{7.8.7}$$

Подставляя (7) в (2), находим

$$s\approx\int_a^b\left[y'(x)+\frac{1}{2y'(x)}\right]dx=$$

$$=\int_a^b y'(x)\,dx+\frac{1}{2}\int_a^b \frac{dx}{y'(x)}=$$

$$=y(b)-y(a)+\frac{1}{2}\int_a^b \frac{dx}{y'(x)}.$$

Итак, мы пришли к следующим приближенным формулам:

$$s\approx(b-a)+\frac{1}{2}\int_a^b (y'(x))^2\,dx,$$

если $|y'(x)|<1$;

$$s\approx y(b)-y(a)+\frac{1}{2}\int_a^b \frac{dx}{y'(x)},\qquad (7.8.8)$$

если $|y'(x)|>1$.

Входящие сюда интегралы проще, чем интеграл в (2), поэтому этими формулами гораздо удобнее пользоваться. Однако формула (2) является точной, а формулы (8) — только приближенными.

Какую же ошибку делаем мы, пользуясь ими? Первая из формул (8) будет тем точнее, чем меньше $|y'|$, а вторая — чем больше $|y'|$; обе формулы дают наихудший результат при $|y'|=1$. Поэтому для оценки погрешности рассмотрим самый невыгодный случай $y'(x)=1$. Ясно, что если $y'(x)=1$, то $y(x)=x+c$; график этой функции — прямая линия.

По точной формуле (2)

$$s=\int_a^b \sqrt{1+1}\,dx=\sqrt{2}\,(b-a)\approx 1{,}415\,(b-a)\qquad (7.8.9)$$

(впрочем, этот последний результат с очевидностью следует и из теоремы Пифагора или элементарной тригонометрии: ведь прямая $y=x+c$ образует с осью абсцисс угол в 45°).

По первой из формул (8):

$$s\approx(b-a)+\frac{1}{2}\int_a^b dx=\frac{3}{2}(b-a)=1{,}5(b-a).\qquad (7.8.10)$$

Вторая формула (8) дает тот же результат: $s\approx 1{,}5\,(b-a)$. Сравнивая (9) и (10), видим, что наибольшая погрешность приближенной формулы близка к 6%.

При вычислении длины дуги кривую следует разбивать на участки, на которых либо всюду $|y'|\leqslant 1$, либо $|y'|\geqslant 1$. Тогда ошибка будет, во всяком случае, не больше 6%. А так как $(y'(x))^2$ принимает значение, равное 1, лишь в отдельных точках кривой, то при правильном разбиении линии на участки ошибка будет меньше 6% (обычно значительно меньше 6%). Длины прямолинейных отрезков (в частности, таких, где $y'(x)\equiv 1$), конечно, находить по приближенной формуле незачем.

Рассмотрим примеры (вычисления почти всюду проведены с точностью до второго знака после запятой).

1. Найти длину дуги *параболы* $y=x^2$ между точками с абсциссами $x=0$ и $x=2$.

Производная $y'=2x$ равна 1 при $x=0{,}5$; и $y'<1$ при $x<0{,}5$, $y'>1$ при $x>0{,}5$. Поэтому длину s_1 дуги, где $0\leqslant x\leqslant 0{,}5$, найдем по первой из формул (8), а длину s_2, где $0{,}5\leqslant x\leqslant 2$, — по второй формуле:

$$s_1\approx(0{,}5-0)+0{,}5\int_0^{0{,}5} 4x^2dx=$$

$$=0{,}5+2\,\frac{(0{,}5)^3}{3}\approx 0{,}58,$$

$$s_2\approx 4-0{,}25+0{,}5\int_{0{,}5}^{2}\frac{dx}{2x}=$$

$$=3{,}75+0{,}25\,(\ln 2-\ln 0{,}5)\approx 4{,}10,$$

значит,

$$s=s_1+s_2\approx 0{,}58+4{,}10=4{,}68.$$

С другой стороны, в силу (2)

$$s=\int_0^2 \sqrt{1+4x^2}\,dx,$$

откуда, сделав замену $2x=z$ и воспользовавшись формулой 33 Приложения II (с. 478), находим точное значение s:

$$\int \sqrt{1+4x^2}\,dx=$$

$$=\frac{1}{2}\left[x\sqrt{4x^2+1}+\frac{1}{2}\ln\left(2x+\sqrt{4x^2+1}\right)\right]+C,\qquad (7.8.11)$$

и, следовательно,

$$s=\frac{1}{2}\left[2\sqrt{17}+\frac{1}{2}\ln\left(4+\sqrt{17}\right)\right]\approx 4{,}65.\qquad (7.8.12)$$

(Недоверчивый читатель может убедиться в справедливости формулы (11), продифференцировав правую часть этой формулы.)

Ошибка при подсчете по формулам (8) составила около 0,7%.

2. Найти длину дуги экспоненциальной кривой $y=e^x$ между точками с абсциссами $x=0$ и $x=1$.

В этом случае $y'=e^x$, и при изменении x от 0 до 1 производная растет от 1 до e. Поэтому используем вторую из формул (8):

$$s\approx e^1-e^0+0{,}5\int_0^1\frac{dx}{e^x}\approx$$

$$\approx 2{,}72-1-0{,}5e^{-x}\Big|_0^1\approx 2{,}04.$$

Точная формула дает для длины дуги значение (см. упр. 1б и 2 ниже)

$$s=\sqrt{1+e^2}-\sqrt{2}+$$

$$+\frac{1}{2}\ln\frac{\sqrt{e^2+1}-1}{\sqrt{e^2+1}+1}-\frac{1}{2}\ln\frac{\sqrt{2}-1}{\sqrt{2}+1}\approx 2{,}00.$$

Ошибка приближенной формулы 2%.

Иногда для приближенного вычисления длины дуги подынтегральную функцию в (2) разлагают в ряд по степеням x. При этом, удерживая надлежащее число членов разложения, можно получить значение длины дуги с любой степенью точности.

Рассмотрим пример. Определим длину окружности. При этом будем искать длину s дуги окружности, соответствующую центральному углу в 30°. Длина окружности $C=12s$. Ясно, что мы получим такой же интеграл, как и в (3), но с другим верхним пределом:

$$s=\int_0^{R/2}\frac{Rdx}{\sqrt{R^2-x^2}} \qquad (7.8.13)$$

$\left(\text{здесь мы используем соотношение } R\sin 30°=\frac{R}{2}\right)$. Подынтегральное выражение преобразуется следующим образом:

$$\frac{R}{\sqrt{R^2-x^2}}=\frac{R}{R\sqrt{1-\left(\frac{x}{R}\right)^2}}=$$

$$=\frac{1}{\sqrt{1-\left(\frac{x}{R}\right)^2}}=\left[1-\left(\frac{x}{R}\right)^2\right]^{-1/2}. \qquad (7.8.14)$$

Выражение (14) разложим в ряд по формуле бинома Ньютона (см. формулу (6.4.3)). Для этого положим $\left(\frac{x}{R}\right)^2=t$. Получим

$$\left[1-\left(\frac{x}{R}\right)^2\right]^{-1/2}=(1-t)^{-1/2}=1+\frac{1}{2}t+$$

$$+\frac{3}{8}t^2+\frac{5}{16}t^3+\frac{35}{128}t^4+\ldots=$$

$$=1+\frac{1}{2}\left(\frac{x}{R}\right)^2+\frac{3}{8}\left(\frac{x}{R}\right)^4+$$

$$+\frac{5}{16}\left(\frac{x}{R}\right)^6+\frac{35}{128}\left(\frac{x}{R}\right)^8+\ldots \qquad (7.8.15)$$

Подставляя (15) в (13) и интегрируя, находим

$$s=R\left(\frac{1}{2}+\frac{1}{6\cdot 2^3}+\frac{3}{40\cdot 2^5}+\right.$$

$$\left.+\frac{5}{16\cdot 7\cdot 2^7}+\frac{35}{128\cdot 9\cdot 2^9}+\ldots\right). \qquad (7.8.16)$$

Члены ряда (16) убывают довольно быстро. Поэтому для получения s достаточно взять несколько первых членов ряда. Взяв один член, получим $s\approx\frac{1}{2}R$, откуда длина всей окружности $C\approx 6R$. Взяв два члена, получим $s\approx 0{,}521R$, т. е. $C\approx 6{,}252R$; три члена ряда дают $s\approx 0{,}523R$, $C\approx 6{,}276R$ и т. д.

Мы знаем, что длина окружности $C=2\pi R$. Сравнивая это с полученными нами результатами, находим такие приближенные значения числа π:

3; 3,126; 3,138, . . .

Чем больше членов ряда (11) взять, тем точнее получим значение π. Значение числа π с четырьмя верными десятичными знаками таково: $\pi\approx 3{,}1416$.

Упражнения

7.8.1. Запишите в виде интеграла длину дуги: а) параболы $y=x^2$ от точки (0, 0) до точки (1, 1); б) экспоненциальной кривой $y=e^x$ от точки $x=0$ до точки $x=1$; в) длину дуги эллипса $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$.

7.8.2. Доведите до конца упр. 1б, выполнив в интеграле замену переменной $1+e^{2x}=z^2$.

7.8.3. Пользуясь приближенными формулами (8), найдите длину дуги цепной линии между точками $x=0$ и $x=2$ $(a=1)$. Сравните полученный результат с точным значением длины дуги.

7.8.4. Найдите приближенно длину дуги гиперболы $xy=-1$ между точками $x=0{,}5$ и $x=1$. [Замечание. В этом случае точного решения получить нельзя, так как интеграл в (2) не может быть выражен с помощью элементарных функций.]

7.8.5. Получите приближенные значения числа π исходя из подсчета длины дуги окружности с центральным углом в 45° (удержите три, четыре и пять членов ряда).

§ 9. Кривизна и соприкасающаяся окружность

С понятием длины дуги связано определение *кривизны* k и *радиуса кривизны* R кривой в некоторой ее точке. Для

окружности радиус кривизны R совпадает с радиусом окружности, а кривизной называется величина $k=\frac{1}{R}$, обратная радиусу. Чем меньше радиус R окружности (т. е. чем больше ее кривизна k), тем круче заворачивается окружность, тем более она «искривлена» (рис. 1). При очень большом R, напротив, искривленность окружности почти не заметна, т. е. окружность обращается в «почти прямую»: так, из-за большой величины радиуса земного шара мы не замечаем искривленности Земли (для того чтобы увидеть это, надо подняться в космос, что пока еще мало кому доступно).

Обратимся теперь к произвольной кривой Г. Ясно, что на рис. 2 линия больше искривлена на участке PQ, где она круто заворачивает налево, чем на участке MN, где ход ее более плавный. Однако какой точный смысл имеет последнее утверждение? Пояснить его можно так.

Рассмотрим угол φ между касательными к кривой в концах M и N дуги MN: если на протяжении этой дуги касательная поворачивалась все время в одном направлении, то угол φ указывает, на сколько именно повернулась на участке MN касательная (или кривая, поскольку направление кривой задается направлением касательной). Для определения степени искривленности кривой нам надо знать еще протяженность s участка (дуги) MN: если s велико, то даже при медленном изменении направления касательной может получиться большое значение φ.

Таким образом, отношение

$$k_{\text{ср}}=\frac{\varphi}{s} \qquad (7.9.1)$$

указывает *среднюю* (или *удельную*!) *кривизну* кривой на дуге MN. *Кривизна* же *в точке* M определяется как п р е д е л средней кривизны при стремлении дуги MN к нулю:

$$k=\lim_{\Delta s\to 0}\frac{\Delta\alpha}{\Delta s},$$

где Δs — длина (малой) дуги MM_1 линии Г, а $\Delta\alpha$ — отвечающее этой дуге приращение угла α, образованного касательной к кривой с осью x, т. е. угол поворота касательной к кривой (угол между касательными t и t_1 к Г в точках M и M_1; рис. 3). Но в силу определения производной (см. § 2.1) последнее равенство можно записать так:

$$k=\frac{d\alpha}{ds}. \qquad (7.9.2)$$

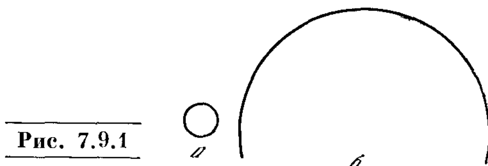


Рис. 7.9.1

Таким образом, кривизну R можно определить как *скорость поворота касательной* к кривой Г при равномерном

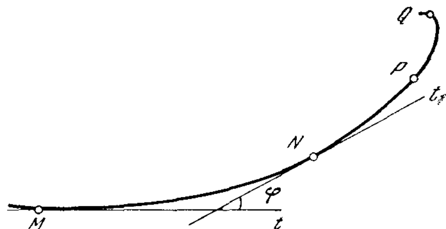


Рис. 7.9.2

движении точки по кривой с единичной (линейной) скоростью, т. е. в условиях, когда точка M проходит путь $MM_1=ds$ за время $dt=ds$.

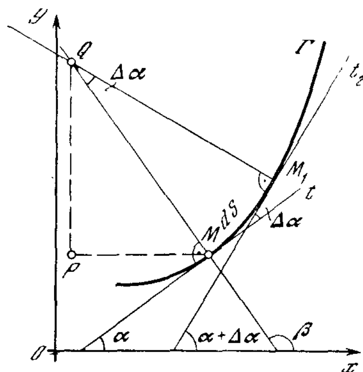


Рис. 7.9.3

Теперь нам остается только вспомнить, что $\operatorname{tg}\alpha=y'$, так что $\alpha=\operatorname{arctg}y'$ и, значит,

$$d\alpha=d(\operatorname{arc\,tg}y')=\frac{dy'}{1+(y')^2}=\frac{y''dx}{1+(y')^2};$$

кроме того (см. § 8), $ds=\sqrt{1+(y')^2}\,dx$. Таким образом, окончательно получаем

$$k=\frac{d\alpha}{ds}=\left[\frac{y''}{1+(y')^2}\,dx\right]\Big/\left[\sqrt{1+(y')^2}\,dx\right]=$$
$$=\frac{y''}{[1+(y')^2]^{3/2}} \qquad (7.9.3)$$

и, значит,

$$R=\frac{1}{k}=\frac{[1+(y')^2]^{3/2}}{y''}. \quad (7.9.3а)$$

Если кривая Г задана параметрическими уравнениями $x=x(t)$, $y=y(t)$ (см. § 1.8), то $\frac{dy}{dx}=\frac{y'}{x'}$ и $\frac{d^2y}{dx^2}=\frac{x'y''-x''y'}{(x')^3}$ (см. (4.4.9)); поэтому из (3) и (3а) следует

$$k=\left[\frac{x'y''-x''y'}{(x')^3}\right]:\left[1+\frac{(y')^2}{(x')^2}\right]^{3/2}=$$
$$=\frac{x'y''-x''y'}{[(x')^2+(y')^2]^{3/2}},$$
$$R=\frac{1}{k}=\frac{[(x')^2+(y')^2]^{3/2}}{x'y''-x''y'}. \quad (7.9.3б)$$

Укажем еще, что поскольку угол α — величина безразмерная (мы используем, как всегда, радианную меру угла), а s имеет размерность $[l]$ длины, то кривизна k кривой имеет размерность $[l^{-1}]$ (скажем, см$^{-1}$), а радиус кривизны, как и полагается радиусу, — размерность длины $[l]$ (например, см).

Понятие радиуса кривизны (а значит, и кривизны) линии можно определить и «более геометрично». Для этого проведем в двух близких точках M и M_1 кривой Г нормали (перпендикуляры к касательным) MQ и M_1Q; через Q мы обозначим точку пересечения нормалей (см. рис. 3). Согласно известной теореме геометрии угол MQM_1 между нормалями равен углу $\Delta\alpha$ между касательными в точках M и M_1 (углы со взаимно перпендикулярными сторонами). Отсюда можно найти расстояние MQ от точки пересечения нормалей до кривой.

Заменим приближенно малый участок кривой дугой окружности. Нормаль к окружности направлена по ее радиусу. Точка пересечения нормалей есть центр окружности. Если бы кривая была окружностью радиуса R, то мы имели бы $ds=Rd\alpha$, или $\frac{d\alpha}{ds}=\frac{1}{R}$; эта величина $\frac{d\alpha}{ds}$ (кривизна окружности) постоянна для любого участка дуги окружности. Подберем теперь дугу окружности так, чтобы она была в о з м о ж н о б о л е е б л и з к о й к дуге MM_1. Слова «возможно более близкой» здесь можно понимать как требование совпадения первых трех членов разложений Тейлора уравнения $y=y(x)$ рассматриваемой кривой

$$y=y(a)+y'(a)(x-a)+$$
$$+\frac{1}{2}y''(a)(x-a)^2+$$
$$+\frac{1}{6}y'''(a)(x-a)^3+\ldots \quad (7.9.4)$$

(мы считаем, что точке M кривой отвечает значение $x=a$ абсциссы) и уравнения $y=Y(x)$ окружности

$$y=Y(a)+Y'(a)(x-a)+$$
$$+\frac{1}{2}Y''(a)(x-a)^2+$$
$$+\frac{1}{6}Y'''(a)(x-a)^3+\ldots \quad (7.9.4а)$$

(см. формулу (6.1.18))[1]. Другими словами, мы требуем совпадения значений не только $y(a)$ и $Y(a)$, но и первых двух производных функций $y(x)$ и $Y(x)$ в точке $x=a$:

$$y(a)=Y(a), \quad y'(a)=Y'(a),$$
$$y''(a)=Y''(a). \quad (7.9.5)$$

Но в силу (3) равенства (5) гарантируют совпадение в точке M кривизны кривой Г с уравнением $y=y(x)$ и окружности Σ с уравнением $y=Y(x)$. А это означает, что радиус $MQ=R$ окружности Σ выражается формулой (3а).

Проходящая через точку M кривой Г окружность Σ, наиболее близкая к Г, называется *соприкасающейся окружностью*[2] этой кривой; радиус R этой окружности и есть *радиус кривизны* кривой, а центр Q окружности Σ называют *центром кривизны* кривой (в точке M). Таким образом, радиус кривизны кривой можно определить как радиус ее соприкасающейся окружности; кривизна же кривой обратна радиусу соприкасающейся окружности кривой.

---

[1] Окружность на плоскости определяется заданием трех чисел (параметров), которыми могут служить, например, коэффициенты a, b и r уравнения $(x-a)^2+(y-b)^2-r^2=0$ окружности. Для нахождения этих чисел (т. е. для определения окружности) необходимо задать т р и условия, которыми могут служить, в частности, выписанные ниже уравнения (5).

[2] Соприкасающуюся окружность Σ кривой Г можно определить еще и следующим образом. Рассмотрим окружность S, проходящую через три точки M_1, M_2 и M_3 кривой Г, близкие к точке M. Когда все эти точки стремятся к точке M, окружность S стремится к соприкасающейся в точке M окружности Σ.

Заметим еще, что поскольку в силу (4), (4а) и (5) разность $Y(x)-y(x)=$ $=\frac{1}{6}[Y'''(a)-y'''(a)]\,dx^3+\ldots$ при переходе через точку M (т. е. при изменении знака dx) меняет знак, то соприкасающаяся окружность Σ в точке соприкосновения с кривой Γ ее пересекает[3]. Это позволяет описать соприкасающуюся окружность так.

Рассмотрим всевозможные окружности σ, касающиеся линии Γ в точке M и имеющие в M то же направление выпуклости, другими словами, проходящие через M, имеющие в M ту же касательную, что и Γ, и такие, что центр σ расположен со стороны вогнутости Γ (рис. 4). Одни из этих окружностей целиком расположены внутри Γ (т. е. со стороны вогнутости Γ); их кривизна будет больше кривизны Γ. Другие, напротив, охватывают Γ снаружи; кривизна таких окружностей меньше кривизны Γ. И лишь одна из окружностей σ переходит в M с одной стороны Γ на другую; эта окружность S_0 теснее всего примыкает к Γ и называется *соприкасающейся окружностью* Σ.

Если условие (5) выполняется уже для прямой (для касательной t к кривой Γ в точке M), то (вычисляемый по формуле (3а)) радиус кривизны Γ обращается в ∞, а кривизна k — в 0; такие точки M называются *точками распрямления* кривой Γ. Ясно, что в точке распрямления линия Γ центра кривизны Q не имеет (он «уходит в бесконечность»).

Из (3) следует, что знаки кривизны k и радиуса кривизны R (до сих пор мы говорили лишь об абсолютной величине k и R, игнорируя знаки этих величин) совпадают со знаком второй производной y'', т. е. характеризуют направление выпуклости кривой (ср. § 2.7). Если кривизна отрицательна, т. е. в точке M имеет место неравенство $y''<0$, то кривая в этой точке расположена ниже касательной в точке M, т. е. она выпукла (или выпукла вверх; рис. 5, *а*); при этом центр кривизны Q расположен ниже точки

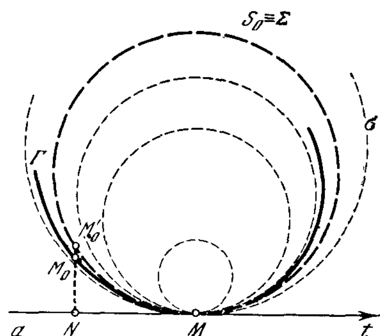


Рис. 7.9.4

M. Если кривизна положительна, т. е. $y''>0$, то кривая расположена выше касательной, т. е. она вогнута (выпукла вниз; рис. 5, *б*); в этом случае центр кривизны Q расположен выше M. Наконец, если кривизна обращается в нуль, т. е. $y''=0$, то, как мы знаем, кривая в точке M меняет направление выпуклости: по одну сторону

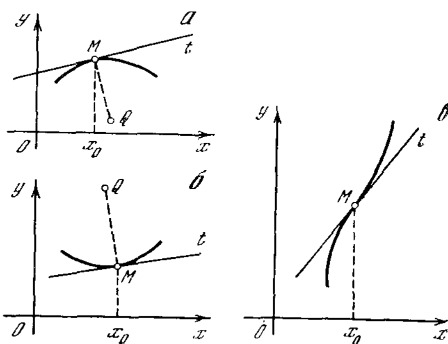


Рис. 7.9.5

от M она направлена выпуклостью вверх, а по другую — выпуклостью вниз (рис. 5, *в*). В такой точке кривая переходит с одной стороны касательной на другую, она меняет направление выпуклости, перегибается, поэтому такие точки и называются точками перегиба. В точке перегиба у уравнения $y=y(x)$ кривой Γ и уравнения $y=Y(x)$ касательной t совпадают не только первые, но и вторые производные: $y''=Y''=0$, т. е. здесь для касательной t выполняются условия

[3] Исключением тут могут служить лишь точки, для которых случайно наряду с условиями (5) имеет место также и равенство $Y'''(a)=y'''(a)$. Примером таких точек могут служить вершины эллипса (точки пересечения эллипса с осями координат), в которых, как это следует из соображений симметрии, соприкасающаяся окружность не может пересечь эллипс; по аналогии с этим случаем все такие точки кривой Γ называются ее *вершинами*.

(5); другими словами, в точке перегиба M роль соприкасающейся окружности Σ играет к а с а т е л ь н а я t, т. е. M есть точка распрямления линии Γ.

Для прямой l, которую ведь тоже можно рассматривать как линию Γ, угол α, образованный касательной к l (это есть сама прямая l) с осью Ox, не меняется; последнее условие означает, что кривизна k прямой тождественно равна нулю (l совсем не искривлена). Для окружности S угол α меняется, но скорость $k=\frac{d\alpha}{ds}$ изменения угла α при движении вдоль окружности сохраняет свое значение; говорят поэтому, что окружности (и прямые) суть линии *постоянной кривизны*. Можно показать что это единственные такие линии.

Пусть $M(x, y(x))$ — точка кривой Γ. Касательная t к этой кривой в точке M имеет наклон y' $(=\operatorname{tg}\alpha)$; нормаль n перпендикулярна t, т. е. образует с осью Ox угол $\beta=\alpha+90°$ (см. рис. 3); ее наклон равен $\operatorname{tg}\beta=-\operatorname{ctg}\alpha=-\frac{1}{y'}$. Поэтому (направленные) проекции MP и PQ на оси координат отрезка MQ длины R соответственно равны $p=R\cos\beta=-R\sin\alpha$ и $q=R\sin\beta=R\cos\alpha$, т. е. координаты ξ, η центра кривизны Q таковы:

$$\xi=x+p=x-R\sin\alpha,$$
$$\eta=y+q=y+R\cos\alpha. \qquad (7.9.6)$$

Учитывая же формулу (3а) для R и то, что, очевидно,

$$\cos\alpha=\frac{1}{\sqrt{1+\operatorname{tg}^2\alpha}}=\frac{1}{\sqrt{1+(y')^2}},$$
$$\sin\alpha=\frac{\operatorname{tg}\alpha}{\sqrt{1+\operatorname{tg}^2\alpha}}=\frac{y'}{\sqrt{1+(y')^2}},$$

окончательно получаем:

$$\xi=x-\frac{\sqrt{[1+(y')^2]^3}}{y''}\,\frac{y'}{\sqrt{1+(y')^2}},$$
$$\eta=y+\frac{\sqrt{[1+(y')^2]^3}}{y''}\,\frac{1}{\sqrt{1+(y')^2}},$$

т. е.

$$\xi=x-\frac{1+(y')^2}{y''}y', \quad \eta=y+\frac{1+(y')^2}{y''}. \qquad (7.9.7)$$

(Это согласуется с условием $\eta>y$ при $k>0$, т. е. при $y''>0$, и $\eta<y$ при $k<0$, т. е. при $y''<0$.)

Если кривая задана параметрическими уравнениями $x=x(t)$, $y=y(t)$, то (ср. (3б))

$$\xi=x-\frac{(x')^2+(y')^2}{x'y''-x''y'}y',$$
$$\eta=y+\frac{(x')^2+(y')^2}{x'y''-x''y'}x' \qquad (7.9.7а)$$

(см. упр. 4).

Множество центров кривизны $Q(\xi, \eta)$ кривой Γ заполняет новую кривую γ, которая называется *эволютой* кривой Γ; сама кривая Γ по отношению к своей эволюте γ называется *эвольвентой* [4].

П р и м е р 1. Парабола $y=x^2$. Здесь $y'=2x$, $y''=2$ $(=\text{const})$, поэтому кривизна k и радиус кривизны R параболы в точке $M(x, y)$ $(=M(x, x^2))$ равны:

$$k=\frac{2}{\sqrt{(1+4x^2)^3}}, \quad R=\frac{\sqrt{(1+4x^2)^3}}{2}.$$

Центр кривизны $Q(\xi, \eta)$ параболы в ее точке $M(x, x^2)$ имеет координаты

$$\xi=x-\frac{1+4x^2}{2}\cdot 2x=-4x^3,$$
$$\eta=x^2+\frac{1+4x^2}{2}=3x^2+\frac{1}{2},$$

т. е.

$$\xi=-4\sqrt{\left(\frac{\eta-1/2}{3}\right)^3}=$$
$$=-\frac{4}{3\sqrt{3}}\sqrt{(\eta-1/2)^3}, \text{ или } \xi=c\eta_1^{3/2}, \qquad (7.9.8)$$

где $c=-\frac{4}{3\sqrt{3}}$, а $\eta_1=\eta-1/2$.

Таким образом, эволютой параболы $y=x^2$ является п о л у к у б и ч е с к а я п а р а б о л а (8), острие которой совпадает с точкой $(0, {}^1/_2)$, а осью симметрии является ось ординат (рис. 6).

П р и м е р 2. Циклоида $x=a(t-\sin t)$, $y=a(1-\cos t)$. Для нее имеем $x'=a(1-\cos t)$, $y'=a\sin t$, $x''=a\sin t$,

[4] Слово э в о л ь в е н т а в переводе с латинского означает «развертывающая», э в о л ю т а — «развернутая», «та, которая развертывается»; эти названия связаны с теми свойствами эволют и эвольвент, о которых рассказано мелким шрифтом в конце настоящего параграфа.

$y''=a\cos t$ (штрихами обозначены производные по параметру t). Поэтому

$$(x')^2+(y')^2=a^2(1-\cos t)^2+a^2\sin^2 t=$$
$$=a^2(2-2\cos t)=4a^2\sin^2\frac{t}{2},$$
$$y''x'-x''y'=a\cos t\cdot a(1-\cos t)-$$
$$-(a\sin t)^2=a^2(\cos t-1)=-2a^2\sin^2\frac{t}{2}.$$

Таким образом, в силу (7а) получаем:

$$\xi=a(t-\sin t)+2a\sin t=a(t+\sin t), \quad (7.9.9)$$
$$\eta=a(1-\cos t)-2a(1-\cos t)=$$
$$=-a+a\cos t.$$

Для того чтобы понять, что это за кривая, введем новый параметр $t_1=t+\pi$ (или $t=t_1-\pi$); так как $\sin t=-\sin t_1$ и $\cos t=-\cos t_1$, то имеем

$$\xi=a(t_1-\sin t_1)-a\pi,$$
$$\eta=a(1-\cos t_1)-2a. \quad (7.9.10)$$

Таким образом, эволюта (9) (или (10)) циклоиды — это такая же ц и к л о и д а $\xi_1=a(t_1-\sin t_1)$, $\eta_1=a(1-\cos t_1)$ (где $\xi_1=\xi+a\pi$, $\eta_1=\eta+2a$), только сдвинутая относительно первоначальной циклоиды на диаметр $2a$ образующего круга вниз и на ширину πa половины арки циклоиды влево (рис. 7).

Для того чтобы разобраться в дальнейших свойствах эволют и эвольвент, продифференцируем равенства (6):

$$d\xi=dx-R\cos\alpha\,d\alpha-\sin\alpha\,dR,$$
$$d\eta=dy-R\sin\alpha\,d\alpha+\cos\alpha\,dR.$$

Но так как (см. (8.2б))

$$dx=\cos\alpha\,ds=\cos\alpha\frac{ds}{d\alpha}d\alpha=R\cos\alpha\,d\alpha,$$
$$dy=\sin\alpha\,ds=\sin\alpha\frac{ds}{d\alpha}d\alpha=R\sin\alpha\,d\alpha,$$

то

$$d\xi=-\sin\alpha\,dR,\quad d\eta=\cos\alpha\,dR. \quad (7.9.11)$$

Поделив равенства (11) одно на другое, получим

$$\frac{d\eta}{d\xi}=-\operatorname{ctg}\alpha=\operatorname{tg}(\alpha+90^\circ). \quad (7.9.12)$$

Отсюда вытекает, что касательная к эволюте γ кривой Г (т. е. к множеству центров кривизны $Q(\xi,\eta)$) образует с осью Ox угол $\beta=\alpha+90^\circ$, т. е. ее направление совпадает с направлением н о р м а л и к Г. А так как точка Q принадлежит нормали n к кривой Г в точке M, то касательная к γ совпадает с прямой n: *касательные к эволюте γ кривой Г совпадают с нормалями*[5] Г (рис. 8).

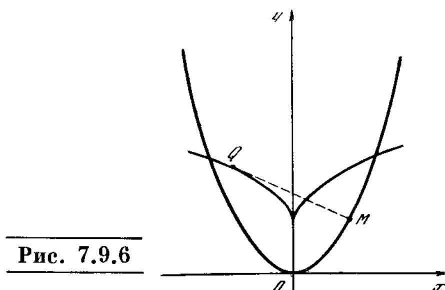


Рис. 7.9.6

С другой стороны, из (11) следует:

$$d\xi^2+d\eta^2=\sin^2\alpha\,dR^2+\cos^2\alpha\,dR^2=dR^2,$$
т. е. $d\sigma=dR,$ (7.9.12а)

где через σ обозначена длина дуги эволюты γ кривой Г (и $d\sigma$, и dR мы считаем положительными).

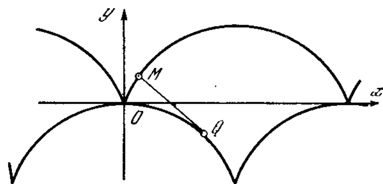


Рис. 7.9.7

Последнее замечание имеет следующий смысл. Рассмотрим такую дугу M_1M_2 кривой Г, что на этой дуге производная $R'=\frac{dR}{ds}$

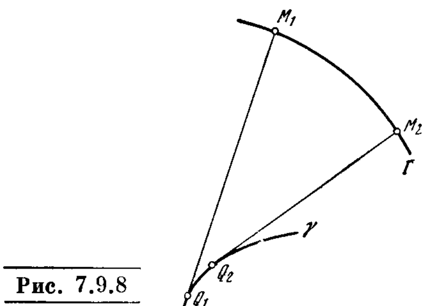


Рис. 7.9.8

сохраняет постоянный знак. При этом мы условимся так выбирать направление отсчета длины дуги s вдоль кривой Г, чтобы эта производная R была п о л о ж и т е л ь н о й; другими словами, будем считать, что дуга отсчитывается вдоль Г в направлении р о с т а ра-

[5] Это выражают, говоря, что эволюта γ кривой Г является *огибающей* нормалей Г.

диуса кривизны R. В таком случае приращение $M_2Q_2 - M_1Q_1 = R_2 - R_1 = \Delta R$ радиуса кривизны вдоль дуги M_1M_2 равно длине $\Delta\sigma$ дуги эволюты. А так как отрезки M_1Q_1 $(=R_1)$ и M_2Q_2 $(=R_2)$ являются касательными к эволюте

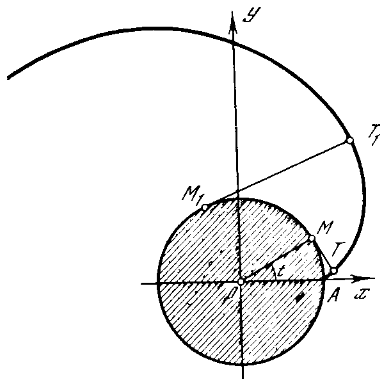


Рис. 7.9.9

γ кривой Γ (см. рис. 8), то процесс образования по линии γ (эволюте) ее эвольвенты Γ можно описать как процесс разматывания нерастяжимой нити, натянутой на «катушку» формы γ: при таком разматывании конец M

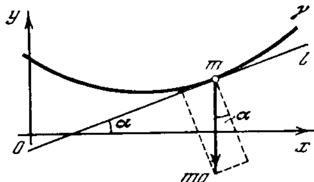


Рис. 7.9.10

нити как раз и будет описывать линию Γ. Именно с этим обстоятельством и связаны названия «эвольвента» и «эволюта» кривой (см. сноску 4 на с. 232).

В качестве примера составим (параметрическое) уравнение э в о л ь в е н т ы о к р у ж-

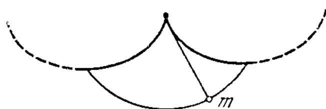


Рис. 7.9.11

н о с т и радиуса r с уравнением $x^2+y^2=r^2$, или $x=r\cos t$, $y=r\sin t$, где t — это угол. Будем считать, что эвольвента проходит через точку A $(r, 0)$ и развертывание намотанной на круглую катушку нити происходит в направлении против часовой стрелки (рис. 9). При этом касательная MT к окружности в точке M $(r\cos t, r\sin t)$ образует с осью Ox угол $\alpha=90°+t$ и отложенный по касательной отрезок MT длины rt (равной длине дуги AM окружности) имеет проекции $rt\cos\alpha = -rt\sin t$ и $rt\sin\alpha = rt\cos t$ на оси координат. Отсюда без труда получаем следующие параметрические выражения координат X, Y точки T (X, Y) (т. е. уравнения эвольвенты окружности):

$$X = r\cos t - rt\sin t,$$
$$Y = r\sin t + rt\cos t. \tag{7.9.13}$$

Учение о эволютах и эвольвентах плоских кривых принадлежит Х. Гюйгенсу; оно было изложено в его книге «О маятниковых часах» (1673) в связи со следующей задачей (ср. [39, с. 96—117]). Составим уравнение движения маятника, т. е. материальной точки массы m, движущейся по линии γ:

$$m\frac{d^2s}{dt^2} = mg\sin\alpha, \tag{7.9.14}$$

где s — длина дуги линии γ, а t — время, так что $\frac{d^2s}{dt^2}$ — это (*тангенциальное*, т. е. направленное по касательной к γ) *ускорение* частицы; g — ускорение свободного падения; α — угол, образованный касательной l к линии γ с горизонталью, и $mg\sin\alpha$ — проекция (вертикальной) силы тяжести mg на l (рис. 10; ср. гл. 9 и 10). При этом период колебания маятника, строго говоря, зависит от размаха, т. е. от той наивысшей точки линии γ, с которой начинается движение маятника. Гюйгенс поставил вопрос о том, при какой форме линии γ период колебаний маятника будет о д н и м и т е м ж е независимо от размаха колебаний. Проанализировав дифференциальное уравнение (14) (что было вовсе не просто: ведь дифференциальное и интегральное исчисление в этот период еще не было создано), Гюйгенс нашел, что для этого необходимо, чтобы масса m перемещалась по *циклоиде*. Но как заставить маятник двигаться по циклоиде? Вот в этой связи Гюйгенс и разработал общую теорию «разверток» (эвольвент) кривых; он обнаружил, что эволютой циклоиды является снова циклоида (ср. с. 233 и рис. 7), и показал, что если расположить у точки подвеса гибкой нити циклоидальные «губы» так, чтобы нить в процессе движения наматывалась на них, то конец нити постоянной длины будет описывать циклоиду (рис. 11). И первые маятниковые часы, сооруженные Гюйгенсом, имели такие «губы»; впрочем, впоследствии Гюйгенс отказался от такой конструкции, выяснив, что практически «губы» не увеличивают точности хода часов, ибо в силу малости углов α в уравнении (14) можно заменить $\sin\alpha$ на α, а из уравнения $m\frac{d^2s}{dt^2} = mg\alpha$ вытекает постоянство

периода колебаний маятника фиксированной длины l (см. § 10.3).

Упражнения

7.9.1. Вычислите кривизну (в произвольной точке кривой): а) эллипса $x=a\cos t$, $y=b\sin t$; б) гиперболы $y=\frac{1}{x}$; в) параболы четвертого порядка $y=ax^4$.

7.9.2. Найдите эволюты: а) эллипса $x=a\cos t$, $y=b\sin t$; б) гиперболы $y=\frac{1}{x}$.

7.9.3. Докажите, что точкам максимума или минимума кривизны линии отвечают «острия» (точки возврата) ее эволюты (ср. с рис. 6 и 7).

7.9.4. Докажите формулы (7а).

§ 10. Стереометрические приложения интегрального исчисления

В § 3.6 была получена формула

$$V=\int_a^b S(x)\,dx, \tag{7.10.1}$$

где $S(x)$ — площадь сечения тела плоскостью $x=\text{const}$, перпендикулярной оси x (советуем читателю повторить вывод этой формулы). При помощи этой формулы там была получена формула для объема пирамиды. Так же получается и формула для объема конуса. Примем за начало координат центр круга, лежащего в основании прямого кругового конуса, а ось x направим по высоте конуса (рис. 1). Пусть $S(x)$ — площадь сечения конуса плоскостью, перпендикулярной к высоте и отстоящей от основания конуса на расстояние x. Это сечение есть круг некоторого радиуса r_x, где

$$\frac{r_x}{r}=\frac{H-x}{H},$$

как это следует из подобия треугольников; здесь r — радиус основания конуса H — его высота. Поэтому $r_x=\frac{r}{H}(H-x)$ и $S(x)=\pi r_x^2=\frac{\pi r^2(H-x)^2}{H^2}$, а, следовательно,

$$V=\int_0^H \pi\frac{r^2}{H^2}(H-x)^2\,dx=$$
$$=-\frac{\pi r^2}{H^2}\frac{(H-x)^3}{3}\Big|_0^H=\frac{\pi r^2H^3}{3H^2}=\frac{\pi r^2H}{3}.$$

Для вычисления объема шара радиуса R совместим начало координат с центром шара (рис. 2). Сечение шара плоскостью, перпендикулярной к оси x и отстоящей от начала координат на

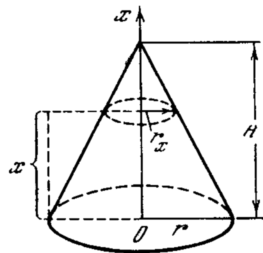


Рис. 7.10.1

расстояние x, есть круг некоторого радиуса R_x, где в силу теоремы Пифагора $R_x=\sqrt{R^2-x^2}$. Поэтому $S(x)=\pi R_x^2=\pi(R^2-x^2)$, и, значит,

$$V=\int_{-R}^{R}\pi(R^2-x^2)\,dx=$$
$$=\pi\left(R^2x-\frac{x^3}{3}\right)\Big|_{-R}^{R}=\frac{4}{3}\pi R^3.$$

Из формулы (1) следует принцип Кавальери[1]: *если два тела, расположенные между двумя параллельными плоскостями P и Q, таковы, что в сечении этих тел любой плоскостью, параллельной P и Q, получаются равновеликие фигуры, то и объемы этих тел*

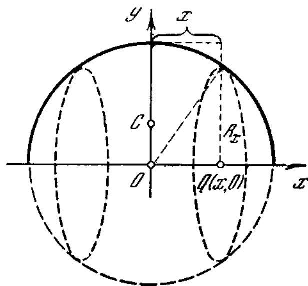


Рис. 7.10.2

равны. В самом деле, при вычислении объемов этих тел по формуле (1), где ось x перпендикулярна плоскостям P и Q, подынтегральная функция $S(x)$ будет для обоих тел одинаковой.

Рассмотрим теперь тело вращения, получающееся из криволинейной тра-

[1] Этот принцип (он сыграл заметную роль в создании концепции интеграла) был сформулирован (по существу без доказательства) в книге «Геометрия неделимых» (1635) Б. Кавальери, ученика Г. Галилея.

пеции, изображенной на рис. 3, при вращении ее вокруг оси x. Сечение полученного тела плоскостью $x=\text{const}$ есть круг радиуса $y=f(x)$; следовательно, здесь $S(x)=\pi y^2$. Пользуясь равенством (1), находим известную формулу для *объема тела вращения*:

$$V=\pi\int_a^b y^2dx. \qquad (7.10.2)$$

Найдем, например, объем *эллипсоида вращения*, получаемого при вращении верхней половины эллипса $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$ вокруг оси x (сделайте чертеж!). Поскольку для эллипса $y=\frac{b}{a}\sqrt{a^2-x^2}$, то из (2) получаем

$$V=\pi\int_{-a}^{a}\frac{b^2}{a^2}(a^2-x^2)\,dx=$$

$$=\frac{\pi b^2}{a^2}\left(a^2x-\frac{x^3}{3}\right)\Big|_{-a}^{a}=\frac{4}{3}\pi ab^2.$$

При $a=b=R$ мы, как и следовало ожидать, получаем объем $\frac{4}{3}\pi R^3$ шара радиуса R.

Выведем теперь формулу для *поверхности* тела вращения (рис. 4). Рассмотрим тело, ограниченное сечениями, отвечающими значениям x и $x+dx$ абсциссы. Обозначим боковую поверхность этого тела через dF. Считая рассматриваемое тело усеченным конусом, получим

$$dF\approx\pi[y(x)+y(x+dx)]\,ds,$$

где ds — длина малого участка кривой, вращением которой образована поверхность тела; при этом $ds=\sqrt{1+(y'(x))^2}\,dx$ (см. § 8). Сумму $y(x)+y(x+dx)$ можно заменить на $2y(x)$, пренебрегая величиной порядка $y'(x)\,dx$, малой по сравнению с $y(x)$ [2]. Поэтому

$$dF\approx 2\pi y(x)\sqrt{1+(y'(x))^2}\,dx.$$

Вся площадь поверхности вращения есть

$$F=2\pi\int_a^b y(x)\sqrt{1+(y'(x))^2}\,dx. \qquad (7.10.3)$$

[2] Заметим, что в выражении для dF сумма $y(x)+y(x+dx)$ умножается на ds, так что величина, которой мы пренебрегаем, имеет порядок малости $dx\,ds\approx dx^2$.

При помощи этой формулы легко находим, например, площадь поверхности шара. Действительно, шар получается вращением верхней полуок-

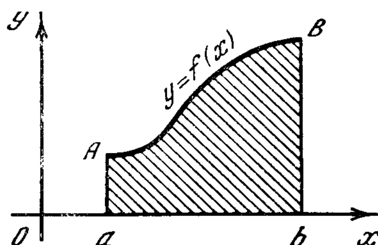


Рис. 7.10.3

ружности вокруг оси x. Уравнение окружности имеет вид $x^2+y^2=R^2$, откуда

$$y=\sqrt{R^2-x^2},\quad y'=-\frac{x}{\sqrt{R^2-x^2}}. \qquad (7.10.4)$$

Подставляя эти значения в (3), получаем

$$F=2\pi\int_{-R}^{R}\sqrt{R^2-x^2}\,\frac{R}{\sqrt{R^2-x^2}}\,dx=$$

$$=2\pi Rx\Big|_{-R}^{R}=4\pi R^2.$$

Более того, из (3) и (4) легко найти также поверхность *шарового слоя* (или

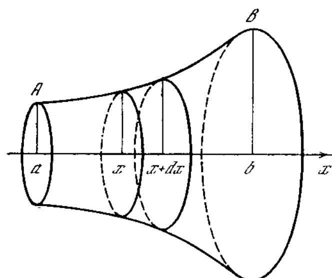


Рис. 7.10.4

шарового пояса), вырезанного из сферы плоскостями $x=a$ и $x=b>a$ (ср. рис. 2; если скажем, $b=R$, то слой обращается в *шаровой сегмент* или *шаровую шапочку*). В самом деле, очевидно, имеем

$$F=2\pi\int_a^b\sqrt{R^2-x^2}\,\frac{Rdx}{\sqrt{R^2-x^2}}=2\pi Rx\Big|_a^b=$$

$$=2\pi R(b-a).$$

Таким образом, искомая поверхность зависит лишь от высоты $b-a$ слоя (сег-

мента) и от радиуса R шара — замечательный по простоте и по красоте результат.

Формулам (2) и (3) можно придать и другую форму.

Пусть изображенная на рис. 3 фигура — это материальная пластинка постоянной толщины. В таком случае масса M этой пластинки равна

$$M = \mu \int_a^b y dx = \mu S, \tag{7.10.5}$$

где μ — плотность материала, из которого сделана пластинка (плотность на единицу площади; см. § 9.13), а $S = \int_a^b y dx$ — площадь пластинки. Расстояние y_C центра тяжести C пластинки от оси Ox выражается формулой

$$y_C = \frac{1}{2M} \int_a^b \mu y^2 dx = \frac{1}{2S} \int_a^b y^2 dx \tag{7.10.6}$$

(ср. с формулой (9.13.12)). Сравнивая формулы (2) и (6), находим

$$V = 2\pi S y_C = 2\pi y_C S \tag{7.10.7}$$

— *объем V тела вращения равен произведению площади S фигуры, вращением которой получается это тело, на длину $2\pi y_C$ окружности, которую описывает при вращении фигуры ее центр тяжести.* (Здесь под центром тяжести плоской фигуры понимается центр тяжести однородной пластинки, имеющей форму этой фигуры.) Эта теорема называется (первой) теоремой Паппа[3].

Формулу (7) и теорему Паппа можно перенести и на случай тела, образованного вращением фигуры произвольной формы — не обязательно криволинейной трапеции.

Рассмотрим, например, тело, образованное вращением вокруг оси Ox изображенной на

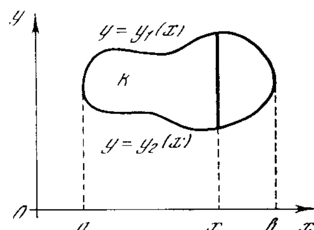


Рис. 7.10.5

рис. 5 фигуры. В этом случае сечением нашего тела перпендикулярной Ox плоскостью будет круговое кольцо, ограниченное окружностями радиусов $y_1(x)$ и $y_2(x)$; поэтому здесь

$$S(x) = \pi [y_1(x)]^2 - \pi [y_2(x)]^2 = \pi (y_1^2 - y_2^2)$$

и, значит, объем тела вращения

$$V = \pi \int_a^b (y_1^2 - y_2^2)\, dx. \tag{7.10.8}$$

С другой стороны, расстояние y_C от оси Ox центра тяжести C однородной пластины, имеющей форму фигуры F, в этом случае равно

$$y_C = \frac{1}{2S} \int_a^b (y_1^2 - y_2^2)\, dx = \frac{\int_a^b (y_1^2 - y_2^2)\, dx}{2\int_a^b (y_1 - y_2)\, dx} \tag{7.10.9}$$

(ср. с формулой (9.13.11)). Поэтому и здесь $V = 2\pi y_C S$.

Аналогично интерпретируется и формула (3). Пусть поверхность тела вращения образована дугой AB (см. рис. 4); представим себе эту дугу как однородную материальную нить постоянной плотности σ (σ — это масса на единицу длины; см. с. 295). В таком случае расстояние y_C центра тяжести C нити от оси Ox будет выражаться формулой

$$y_C = \frac{1}{M} \int_{s_0}^{s_1} \sigma y(x)\, ds = \frac{1}{L} \int_{s_0}^{s_1} y(x)\, ds =$$

$$= \frac{1}{L} \int_a^b y(x) \sqrt{1 + (y'(x))^2}\, dx; \tag{7.10.10}$$

здесь $x(s_0) = a$, $x(s_1) = b$, ds — элемент длины дуги линии AB, $L = \int_{s_0}^{s_1} ds =$

---

[3] Папп Александрийский (2-я половина III в. н. э.) — последний из великих греческих математиков; в его сочинениях излагались (чаще всего без доказательства) многие результаты, как принадлежащие его предшественникам, так и полученные им самим. В литературе (первая и вторая) теоремы Паппа часто называются теоремами Паппа — Гульдина или даже просто теоремами Гульдина, что, однако, совершенно неосновательно: формулировки обеих теорем имелись еще у Паппа, а их доказательства, предложенные швейцарским монахом и математиком-любителем Паулем Гульдином (1577—1643), были весьма сомнительны и значительно уступали доказательствам тех же теорем, данным Б. Кавальери и знаменитым И. Кеплером (последние доказательства Гульдин ожесточенно и несправедливо критиковал).

$= \int_a^b \sqrt{1+(y'(x))^2}\,dx$ — длина нити, $M=\sigma L$ — ее масса (ср. с формулами (9.13.4) и (9.13.5б)).

Сравнивая выражения (3) и (10), находим

$$F = 2\pi L y_C = 2\pi y_C L \qquad (7.10.11)$$

— *поверхность тела вращения равна произведению длины L дуги, при вращении которой получается эта поверхность, на длину $2\pi y_C$ окружности, описываемой при вращении дуги ее центром тяжести.* И этот результат (вторая теорема Паппа) переносится на случай поверхности вращения, образованной вращением линии произвольной формы

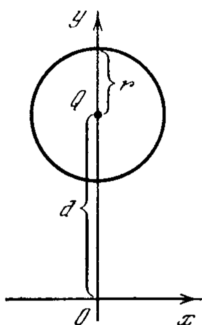
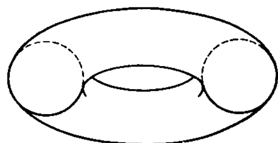


Рис. 7.10.6

Рис. 7.10.7



(например, той, которая ограничивает фигуру K на рис. 5), а не только линии AB (см. рис. 3), ограничивающей криволинейную трапецию.

Теоремы Паппа (7) и (11) часто бывают полезны. Определим, например, объем и поверхность «баранки» (тора), образованной вращением вокруг оси Ox круга радиуса r, центр которого удален от оси вращения на расстояние d (рис. 6). Здесь, очевидно, центр тяжести C круга (и ограничивающей круг окружности) совпадает с (геометрическим) центром Q нашего круга. Поэтому объем баранки (рис. 7) равен

$$V = 2\pi d \cdot \pi r^2 = 2\pi^2 r^2 d, \qquad (7.10.12)$$

а ее поверхность

$$F = 2\pi d \cdot 2\pi r = 4\pi^2 r d. \qquad (7.10.13)$$

Теоремы Паппа могут быть использованы и «в обратном направлении». Шар (см. рис. 2) можно получить вращением полукруга вокруг ограничивающего его диаметра. Поэтому если радиус шара равен R, то его объем

$$V = 2\pi y_C \frac{\pi R^2}{2} = \pi^2 R^2 y_C.$$

Зная, что объем V шара равен $^4/_3\,\pi R^3$, находим отсюда расстояние y_C центра тяжести полукруга от его диаметра:

$$y_C = \frac{4}{3}\pi R^3/(\pi^2 R^2) = \frac{4R}{3\pi} \approx 0{,}4R.$$

Аналогично с помощью второй теоремы Паппа может быть найден и центр тяжести полуокружности (см. упр. 3).

Упражнения

7.10.1. Найдите объем конуса, пользуясь тем, что конус — это тело, получаемое при вращении прямоугольного треугольника вокруг одного из катетов.

7.10.2. Найдите объем тела, полученного при вращении вокруг оси Ox фигуры, ограниченной сверху линией $y=\sqrt{x}$, снизу — осью x, справа — вертикальной прямой $x=2$.

7.10.3. Воспользовавшись второй теоремой Паппа, найдите центр тяжести полуокружности.

§ 11. Как строить кривые

Самый примитивный способ построить график функции $f(x)$ — построение «по точкам»: оно состоит в вычислении $f(x_n)$ для большого числа значений x_n, которые чаще всего выбирают в виде $x_n=x_0+na$; $n=0;\ \pm1,\ \pm2, \ldots$ Ясно, что такой способ не экономичен. Для того чтобы проследить за изменением функции на интервале Δx, нужно выбрать шаг a гораздо меньшим, чем Δx: $a \ll \Delta x$. Но при малом шаге потребуется очень большое число значений x_n для того, чтобы охватить всю интересующую нас область. А между тем с задачей построения графиков мы встречаемся весьма часто: ведь знание графика доставляет нам ряд важных сведений о функции, например, позволяет установить число вещественных корней уравнения $f(x)=0$, найти промежутки, в которых располагаются корни, получить информацию о максимумах и минимумах функции и т. д.

Приемы, рассмотренные в § 1, 2, позволяют гораздо быстрее и надежнее построить график, составить общее представление о виде кривой. Для этого прежде всего следует найти характерные точки графика — максимумы, минимумы, точки разрыва или изломов, точки перегиба и т. д.

Проиллюстрируем это на примере построения графика **многочлена** третьей степени

$$y = ax^3 + bx^2 + cx + d. \qquad (7.11.1)$$

Построим, например, график функции

$$y = 0{,}5x^3 - 0{,}75x^2 - 3x + 2{,}5. \qquad (7.11.2)$$

Найдем прежде всего **максимумы** и **минимумы**. Приравнивая нулю производную функции (2), получим

$$y' = 1{,}5x^2 - 1{,}5x - 3 = 0, \qquad (7.11.3)$$

откуда находим два корня уравнения (3): $x_1 = -1$, $x_2 = 2$.

Исследуем каждое из этих значений в отдельности. Для этого найдем y'': $y'' = 3x - 1{,}5$; $y''(-1) = -4{,}5 < 0$, $y''(2) = 6 - 1{,}5 = 4{,}5 > 0$. Значит, при $x = -1$ функция y имеет **максимум**:

$$y_{\max} = -0{,}5 - 0{,}75 + 3 + 2{,}5 = 4{,}25.$$

Напротив, при $x = 2$ функция имеет **минимум**:

$$y_{\min} = -2{,}5.$$

Теперь посмотрим, как будет вести себя многочлен y при **очень больших** по абсолютной величине значениях x. При больших x член, содержащий x^3, будет значительно превосходить по абсолютной величине остальные слагаемые правой части (2). Поэтому при больших по абсолютной величине x знак многочлена (2) определяется знаком выражения $0{,}5x^3$: при $x \gg 0$ и $y \gg 0$, т. е. правая ветвь графика уходит вверх; при $x \ll 0$ и $y \ll 0$, т. е. левая ветвь графика уходит вниз. (Ясно, что так же будет обстоять дело и в случае «общего» многочлена (1), если только $a > 0$; напротив, при $a < 0$ левая ветвь графика уходит вверх, а правая — вниз.)

Найдем теперь **точки перегиба**, т. е. точки, в которых $f''(x) = 0$ (см. § 9). Из (3) следует $y'' = 3x - 1{,}5$, т. е. $y'' = 0$ лишь при $x = 0{,}5$. Отметим, что график многочлена третьей степени всегда имеет точку перегиба (и притом единственную), в которой касательная к графику его пересекает (рис. 1). Действительно, если y есть многочлен третьей степени, то уравнение $y'' = 0$ является уравнением первой степени. Оно всегда имеет единственный корень x_0. В этой точке разность $y - \tilde{y}$, где $\tilde{y}$ — ордината касательной к графику, имеет вид $A(x - x_0)^3$, т. е. меняет знак при переходе x через значение x_0; поэтому здесь касательная наверняка график пересекает.

Вернемся к построению нашего графика. Вычислим ординату y точки перегиба; получим $y = 0{,}875$. Определим

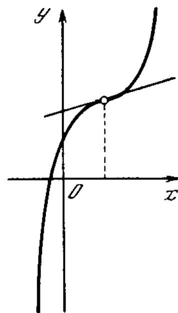


Рис. 7.11.1

еще направление касательной к графику в точке перегиба. Из уравнения (3) получаем $\operatorname{tg} \alpha = y'(0{,}5) = 3{,}375$. Используя все приведенные выше соображения, мы приходим к форме кривой (2), изображенной на рис. 2.

Конечно, если не вычислять больше никаких значений функции, то получим график, дающий грубую качественную картину поведения функции.

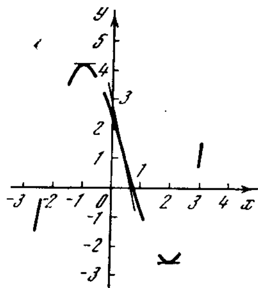


Рис. 7.11.2

Однако даже такой график дает возможность подсчитать число корней функции (т. е. число точек пересечения графика функции с осью x) и грубо оценить их значения. В случае рис. 2 видим, что корней три: один из них лежит где-то между 0,5 и 2, другой корень обязательно положителен (он даже больше 2), а третий — отрицателен (он меньше —1).

График можно уточнить, если подсчитать значения функции еще при нескольких значениях x. Найдем, например, еще три значения нашей функции, При $x = 0$ имеем $y = 2{,}5$; это позволит

нам лучше представить ход кривой между максимумом и минимумом. При $x=3$ получаем $y=0{,}25$; это значение дает представление о скорости подъема правой ветви кривой. Аналогично, чтобы как-то оценить скорость спада левой ветви кривой, найдем $y(-2)=1{,}5$. Используя эти значения, получаем график, изображенный на рис. 3.

По такому графику можем сделать более точные заключения о корнях: один из них лежит между $x=0{,}5$ и $x=1$; второй — между $x=2$ и $x=3$, ближе к $x=3$; третий корень меньше $x=-2$ (вероятно, его значение близко к $x=-2{,}5$).

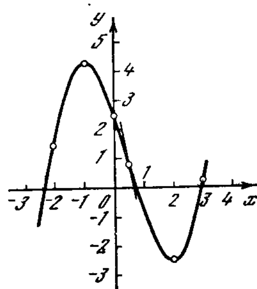


Рис. 7.11.3

Может случиться, что, приравняв производную нулю, мы не получим вещественных корней. Это означает, что исследуемый многочлен не имеет ни максимума, ни минимума. Так как все сказанное о поведении многочлена третьей степени при очень больших по абсолютной величине значениях x остается в силе, то в этом случае график пересекает ось x только в одной точке (многочлен имеет один вещественный корень, ср. рис. 1).

Наконец, производная многочлена может иметь лишь один (двойной) корень x_0. В этом случае производная имеет вид

$$y'=A(x-x_0)^2, \tag{7.11.4}$$

откуда, интегрируя, получаем

$$y=\frac{A}{3}(x-x_0)^3+C. \tag{7.11.5}$$

Из (5) следует что здесь многочлен (1) отличается от полного куба лишь постоянным слагаемым. Ясно, что в этом случае ни максимума, ни минимума функция $y(x)$ не имеет (см. § 1, пример 1а). График функции пересекает ось x в одной точке, которую можно найти, приравняв y нулю:

$$\frac{A}{3}(x-x_0)^3+C=0,$$

т. е.

$$(x-x_0)^3=-\frac{3C}{A}, \quad x=x_0-\sqrt[3]{\frac{3C}{A}}. \tag{7.11.6}$$

Нахождение максимума и минимума многочлена третьей степени (1), а значит, и исследование его графика всегда можно довести до конца, потому что, приравняв производную нулю, получим **квадратное** уравнение, корни которого найти нетрудно. (Однако, разумеется, в общем случае дело вовсе не всегда будет обстоять столь благополучно.) Кроме того, многочлен (любой степени) хорош тем, что он определен при всех значениях x и его график не имеет ни разрывов, ни изломов; при x, неограниченно возрастающем по абсолютной величине, абсолютная величина y также возрастает неограниченно (причем тем быстрее, чем выше степень многочлена). Но при произвольной (хотя бы алгебраической) функции y ситуация может оказаться иной.

В общем случае построение графика требует изучения поведения соответствующей функции «на бесконечности», т. е. изучения характера ее изменения при $x\to\infty$ и при $x\to-\infty$: если, скажем, при неограниченно возрастающем x значение функции все меньше отличается от фиксированного числа C, то функция имеет *горизонтальную асимптоту* $y=C$, т. е. при неограниченном росте x ее график «прижимается» к прямой $y=C$. Аналогично если при $x=a$ функция претерпевает разрыв и, например, при приближении к значению $x=a$ значения y неограниченно возрастают, то прямая $x=a$ является *вертикальной асимптотой* графика функции $y=f(x)$, к которой этот график «неограниченно прижимается» при приближении x к значению a. Ясно также, что построение графика функции естественно начинать с нахождения области ее определения; кроме того, надо выяснить, не является ли функция **четной** или **нечетной** (см. § 1.7), ибо в случае, скажем, четной функции нам достаточно построить по-

ловину графика, отвечающую положительным значениям x, а затем продолжить этот график по симметрии.

Проиллюстрируем сказанное несколькими простыми примерами. Пусть

$$y = \frac{x-1}{x+1}. \qquad (7.11.7)$$

Ясно, что эта функция определена при всех x, кроме значения $x = -1$, при котором знаменатель стоящей в правой части дроби обращается в нуль: при $x = -1$ функция претерпевает разрыв. Так как при приближении значения x к величине -1 значения y по абсолютной величине неограниченно возрастают, то прямая l с уравнением $x = -1$ является *вертикальной асимптотой* графика функции. Нетрудно также видеть, что вблизи значения $x = -1$ величина y будет положительна слева от прямой l и отрицательна справа от нее (рис. 4).

Дальнейшие детали поведения функции связаны с тем, что значения

$$y = 1 - \frac{2}{x+1} \qquad (7.11.8)$$

при больших по абсолютной величине значениях x будут сколь угодно близки к 1 $\left(\text{ибо дробь } \frac{2}{x+1} \text{ здесь становится}\right.$ весьма малой); поэтому прямая $y=1$ будет служить *горизонтальной асимптотой* графика функции. Наконец, в силу (8)

$$y' = \frac{2}{(x+1)^2}, \quad y'' = -\frac{4}{(x+1)^3},$$

т. е. $y' > 0$ при всех значениях x, при которых функция определена; $y'' < 0$ при $x > -1$ и $y'' > 0$ при $x < -1$. Таким образом, наша функция при всех x (кроме $x = -1$, где она вообще не определена) *возрастает*; при $x > -1$ график функции обращен *выпуклостью вверх*, а при $x < -1$ — *выпуклостью вниз*; точек перегиба, в которых имело бы место равенство $y'' = 0$, график не имеет. Кроме того, полезно еще заметить, что $y = 0$ при $x = 1$, и только при этом значении x. График функции пересекает ось x в точке $y = y(0) = -1$ (см. рис. 4).

Вот два несколько более сложных примера. Если

$$y = \frac{x^2-2}{x^2-1}\left(= 1 - \frac{1}{x^2-1}\right), \qquad (7.11.9)$$

то исследование функции упрощается ее *четностью*: график функции симметричен относительно оси Oy. Далее, наша функция определена при всех x, кроме $x=1$ и $x=-1$; точки же 1 и -1 суть точки обращения функции в бесконечность. Таким образом, прямые $x=1$

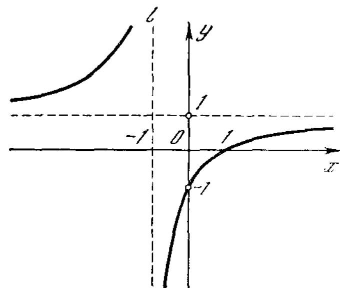


Рис. 7.11.4

и $x = -1$ являются *вертикальными асимптотами* графика функции; прямая же $y=1$ является его *горизонтальной асимптотой* (см. вторую форму записи функции (9). Ясно также, что уравнение $y(x) = 0$ имеет корни $x = \pm\sqrt{2}$, т. е. $x \approx 1{,}41$ и $x \approx -1{,}41$.

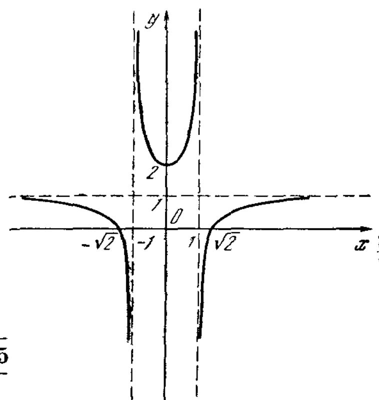


Рис. 7.11.5

Очевидно (тут удобнее вторая форма записи функции (9)),

$$y' = \frac{2x}{(x^2-1)^2},$$

$$y'' = \frac{2}{(x^2-1)^2} - \frac{2 \cdot 2x \cdot 2x}{(x^2-1)^3} = -\frac{6x^2+2}{(x^2-1)^3}.$$

Таким образом, производная y' обращается в нуль лишь при $x=0$; так как $y''(0) = 2 > 0$, то значению $x=0$ отвечает (локальный) минимум функции; при этом $y(0) = 2$. При $x < 0$ функция убывает, а при $x > 0$ — растет; точек перегиба она не имеет, ибо равенство $y''(x) = 0$ невозможно; при $x^2 < 1$, т. е.

при $-1 < x < 1$, вторая производная положительна, т. е. функция обращена выпуклостью вниз; при $|x| > 1$ функция обращена выпуклостью вверх.

График функции изображен на рис. 5.

Наконец, для функции

$$y = \sqrt{\frac{x^2-2}{x^2-1}} = \left(\frac{x^2-2}{x^2-1}\right)^{1/2} \left(= \left(1 - \frac{1}{x^2-1}\right)^{1/2}\right) \qquad (7.11.10)$$

область определения совпадает с областью положительности функции (9),

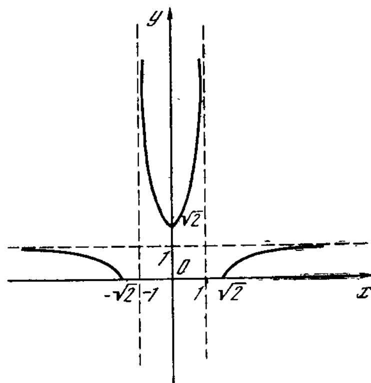


Рис. 7.11.6

т. е. функция определена лишь при $|x| < 1$ и при $|x| > \sqrt{2} \approx 1,4$. Поскольку значения этой функции получаются извлечением корня квадратного из соответствующих значений функции, график которой изображен на рис. 5, мы можем сразу начертить требуемый график (рис. 6). Впрочем, при желании получить более полное представление о поведении интересующей нас кривой можно вычислить несколько значений $y(x)$, отвечающих тем или иным точкам кривой, или значения $y'(x)$, указывающие угол наклона касательной к графику в точке $(x, y(x))$ (так, например, для кривых рис. 5 и 6 полезно найти наклон $y'(x)$ касательной в точке (1,41; 0)).

Разумеется, детальность, с которой строим мы ту или иную кривую, зависит от тех причин, по которым она нас заинтересовала; в ряде случаев вполне достаточно получить грубое качественное представление поведения линии, в других требуется составить о ней более полное представление.

Конечно, изложенное в настоящем параграфе касается лишь «школьных» методов построения линий, заданных своими уравнениями. Совсем другие возможности представляет для этого современная вычислительная техника, включающая даже машины, которые сами вычерчивают линию на экране осциллографа (на дисплее машины) по введенному в память машины уравнению (или программе нахождения точек (x, y) кривой).

Упражнения

7.11.1. Найдите максимумы и минимумы следующих функций и постройте их графики: а) $y = x^3 - 3x^2 + 2$; б) $y = x^3 - 3x^2 + 3x - 15$; в) $y = x^3 - 3x^2 + 6x + 3$.

7.11.2. Определите число вещественных корней уравнений: а) $2x^3 - 3x^2 - 12x + 15 = 0$; б) $4x^3 + 15x^2 - 18x - 2 = 0$; в) $2x^3 - x^2 - 4x + 3 = 0$; г) $x^3 - x^2 + 2 = 0$.

7.11.3. Постройте графики функций: а) $y = \frac{x^3}{x-1}$; б) $y^2 = x^3(1-x)$; в) $y^2 = x^4(1+x)$; г) $y^3 = x^2(1-x)^2$.

ЧАСТЬ ВТОРАЯ

ПРИЛОЖЕНИЯ ВЫСШЕЙ МАТЕМАТИКИ К НЕКОТОРЫМ ВОПРОСАМ ФИЗИКИ И ТЕХНИКИ

Глава 8

РАДИОАКТИВНЫЙ РАСПАД И ДЕЛЕНИЕ ЯДЕР

§ 1. Основные характеристики радиоактивного распада

Установленный на опыте *основной закон радиоактивного распада* состоит в том, что отношение числа распавшихся за единицу времени атомов к общему числу атомов является постоянной величиной, зависящей только от вида атома. При этом подразумевается, что общее число атомов весьма велико.

Указанное отношение называется *вероятностью распада*. Обозначим количество атомов, которое еще не распалось к моменту времени t, через $N(t)$. В момент времени $t+dt$ нераспавшихся атомов будет $N(t+dt)$. Поэтому за время dt (от момента t до момента $t+dt$) распадается $N(t)-N(t+dt)\approx -dN$ атомов. Разделив отношение числа $-dN$ распавшихся за время dt атомов к общему числа N атомов $\left(\text{т. е. долю } -\frac{dN}{N} \text{ распавшихся атомов}\right)$ на время dt, мы получим вероятность распада $\omega=-\frac{dN}{Ndt}$. Отсюда следует, что

$$\frac{dN}{dt}=-\omega N. \tag{8.1.1}$$

Из этого соотношения, поскольку размерность $\frac{dN}{dt}$ совпадает с размерностью отношения $\frac{N}{t}$, следует, что вероятность распада ω имеет размерность 1/с [1].

Начальное условие состоит в задании числа атомов в начальный момент времени: $N=N_0$ при $t=t_0$.

Решая уравнение (1) способом, изложенным в § 6.6 (см. равенства (6.6.9) и (6.6.10)), и пользуясь начальным условием, находим

$$N(t)=N_0e^{-\omega t} \tag{8.1.2}$$

(советуем читателю проделать все выкладки). Однако в случае простейшего уравнения (1), т. е. когда производная пропорциональна искомой функции, можно предложить более простой способ решения уравнения.

В гл. 4 и 6 мы выяснили, что производная от показательной функции пропорциональна самой функции:

$$\frac{d\,(a^x)}{dx}=\text{const}\cdot a^x;$$

в частности

$$\frac{d\,(Ce^{kx})}{dx}=Cke^{kx},$$

если C и k постоянны. Исходя из этого свойства показательной функции, предположим, что решение уравнения (1) имеет вид

$$N=Ce^{kt}, \tag{8.1.3}$$

и постараемся подобрать C и k так, чтобы удовлетворялись и уравнение,

---

[1] Следовательно, вероятность здесь понимается не так, как в утверждении о том, что при бросании монеты вероятность выпадения герба равна половине. Определение вероятности распада как отношения числа распадов в единицу времени к начальному числу атомов справедливо только в том случае, если число распадов в единицу времени (например, в секунду) составляет малую долю общего числа атомов. Точное определение вероятности распада дается именно формулой $\omega=-\frac{1}{N}\frac{dN}{dt}$, т. е. вероятность распада, п о о п р е д е л е н и ю, равна отношению числа распадов за малый промежуток времени к общему числу атомов и к величине промежутка времени.

и начальное условие. Дифференцируя (3), получим

$$\frac{dN}{dt} = Cke^{kt} = kN.$$

Теперь подставим это в (1): $kN = -\omega N$, откуда $k = -\omega$. Полагая в (3) $t=0$ и пользуясь начальным условием, получаем $C = N_0$. Итак, $N = N_0 e^{-\omega t}$, где величина $-\omega t$, стоящая в показателе степени, как и должно быть, является безразмерной.

Радиоактивные атомы обычно характеризуют их *периодом полураспада* T, который представляет собой время, в течение которого число атомов N вследствие распада уменьшается в двое по сравнению с начальным. Определим период полураспада T. В силу (2) $N(T) = N_0 e^{-\omega T}$; с другой стороны, по определению числа T, имеем $N(T) = \frac{1}{2} N_0$. Поэтому

$$N_0 e^{-\omega T} = \frac{1}{2} N_0, \quad e^{-\omega T} = \frac{1}{2},$$

т. е.

$$-\omega T = -\ln 2, \quad T = \frac{\ln 2}{\omega} \approx \frac{0{,}69}{\omega}. \quad (8.1.4)$$

Следовательно, период полураспада обратно пропорционален вероятности распада.

Каждый атом, прежде чем распасться, существует некоторое время, которое называется *временем жизни* атома.

Найдем среднее время жизни $\bar{t}$ атома данного радиоактивного вещества. Пусть в начальный момент $t=0$ (можно считать, что момент $t=0$ характеризует возникновение радиоактивного элемента) у нас имелось N_0 атомов. За время от t до $t+dt$ распадается примерно

$$-dN = \omega N dt$$

атомов. Все атомы этой группы прожили одинаково долго, их время жизни есть t. Среди взятых в начальный момент $t=0$ атомов имеются группы атомов, которым предстоит прожить неодинаковое время от общего для всех атомов момента изготовления вещества до различного для разных атомов момента распада. Чтобы найти среднее время жизни, надо умножить время жизни каждой группы на число атомов в этой группе, сложить эти величины для всех групп и поделить на общее число атомов во всех группах.

Так как нам придется складывать весьма большое количество очень малых слагаемых, то вместо суммы появится интеграл, поэтому (ср. с § 7.7)

$$\bar{t} = \frac{\int_0^\infty t\omega N dt}{\int_0^\infty \omega N dt}. \quad (8.1.5)$$

Подставим в (5) выражение (2) для N. Знаменатель дроби в правой части (5) равен

$$\int_0^\infty \omega N dt = \int_0^\infty \omega N_0 e^{-\omega t} dt = \omega N_0 \int_0^\infty e^{-\omega t} dt =$$
$$= -\omega N_0 \frac{e^{-\omega t}}{\omega}\Big|_0^\infty = N_0,$$

как и следовало ожидать, так как интеграл в знаменателе указывает общее число всех распавшихся атомов, равное, очевидно, числу атомов в начальный момент времени.

Интеграл, стоящий в числителе правой части (5), вычисляется методом интегрирования по частям. Полагаем $t = f$, $e^{-\omega t}dt = dg$, $-\frac{1}{\omega} e^{-\omega t} = g$. При этом получаем

$$\omega N_0 \int_0^\infty t e^{-\omega t} dt =$$
$$= \omega N_0 \left(-\frac{1}{\omega} t e^{-\omega t} + \int \frac{1}{\omega} e^{-\omega t} dt\right)\Big|_0^\infty =$$
$$= \omega N_0 \left(-\frac{1}{\omega} t e^{-\omega t} - \frac{1}{\omega^2} e^{-\omega t}\right)\Big|_0^\infty = \frac{N_0}{\omega}.$$

Из формулы (5) теперь следует

$$\bar{t} = \frac{N_0}{\omega N_0} = \frac{1}{\omega} \quad (8.1.6)$$

— среднее время жизни атома в точности равно величине, обратной вероятности распада. Пользуясь этим, можно основное уравнение (1) и его решение (2) записать так:

$$\frac{dN}{dt} = -\frac{N}{\bar{t}}, \quad (8.1.7)$$

$$N = N_0 e^{-t/\bar{t}}. \quad (8.1.8)$$

При этом надо помнить, что время t есть независимая переменная, от кото-

рой зависит число атомов N. Величина же $\bar{t}$ есть постоянная, характеризующая данный тип радиоактивных атомов.

Из (8) следует, что за время $t=\bar{t}$ число атомов уменьшается от N_0 до $N_0 e^{-1}=\frac{N_0}{e}$, т. е. приблизительно в 2,72 раза.

В силу (7) начальная скорость распада такова, что если бы число атомов, распадающихся за единицу времени, не уменьшалось, то все атомы распались бы за время $\bar{t}$. Действительно, при $t=0$ было N_0 атомов и скорость распада была равна $\left.\frac{dN}{dt}\right|_{t=0}=-\frac{N_0}{\bar{t}}$. При такой скорости для полного распада нужно время, равное $\bar{t}$. Из (4) следует, что $\omega=\frac{\ln 2}{T}$, поэтому

$$\bar{t}=\frac{T}{\ln 2}\approx 1{,}45T.$$

Величиной $\bar{t}$ в расчетах пользоваться удобнее, чем периодом полураспада T.

Упражнения

8.1.1. Среднее время жизни радия 2400 лет. Определите период полураспада радия.

8.1.2. Вначале имелось 200 г радия. Сколько его останется через 300 лет?

8.1.3. За 500 лет распалось 10 г радия. Сколько его было в начальный момент?

8.1.4. Определите, через сколько времени распадается 1%, 10%, 90%, 99% от первоначального запаса радия.

8.1.5. Содержание радия на Земле в различных породах (отношение числа атомов радия к числу атомов породы) в среднем близко к 10^{-12}. Каково было содержание радия 10 000 лет назад; 10^6 лет назад; $5\cdot 10^9$ лет назад ($5\cdot 10^9$ лет — возраст Земли)?

§ 2. Измерение среднего времени жизни радиоактивных атомов

Среднее время жизни $\bar{t}$ различных радиоактивных атомов весьма различно. Так, например, известно несколько изотопов у р а н а. Один из них ^{238}U — уран с атомной массой 238 — имеет среднее время жизни $\bar{t}=7\cdot 10^9$ лет. Другой изотоп ^{235}U имеет среднее время жизни $\bar{t}=10^9$ лет (получение атомной энергии на атомных электростанциях при делении урана происходит в основном за счет ^{235}U). Среднее время жизни р а д и я 2400 лет [1].

Однако не надо думать, что среднее время жизни всех радиоактивных атомов исчисляется тысячелетиями. Среди радиоактивных веществ, встречающихся в природе и изученных еще супругами Кюри и Эрнестом Резерфордом, имеется п о л о н и й со средним временем жизни около 200 дней, р а д и й A со средним временем жизни 4 мин и р а д и й C′ со средним временем жизни $2\cdot 10^{-4}$ с.

В последние десятилетия в связи с развитием ядерной физики и использованием атомной энергии было открыто огромное количество (свыше 1000) различных радиоактивных веществ с самым различным средним временем жизни.

Если в момент времени t имеется $N(t)$ нераспавшихся атомов, то в единицу времени распадается $n(t)=\omega N(t)=-\frac{dN}{dt}$ атомов. Величина $n(t)\left(=n_0 e^{-\omega t}\right.$, где $n_0=-\left.\frac{dN}{dt}\right|_{t=t_0}$; ср. с (1.2)$\Big)$ есть с к о р о с т ь р а с п а д а атомов (см. § 1).

Предположим, что мы наблюдаем за распадом урана 10 лет. За это время в 1 г распадается около $4\cdot 10^{12}$ атомов. Обнаружить, что вместо $2{,}5\cdot 10^{21}$ атомов осталось $2{,}5\cdot 10^{21}-4\cdot 10^{12}$ атомов было бы очень трудно. Измерение *числа распадов* часто оказывается более легким делом, чем измерение количества нераспавшихся атомов.

Проводя опыты с радиоактивными веществами, имеющими не очень большое среднее время жизни (от нескольких минут до нескольких дней), удается с большой точностью проверить формулу $n=n_0 e^{-\omega t}$ (вытекающую из пропорциональности n и N), а тем самым подтвердить и формулы (1.1) и (1.2). Для этого можно поступить следующим образом. Будем подсчитывать число распадов за небольшие промежутки времени. Поделив число распадов на длину промежутка времени, получим скорость распада в различные моменты времени.

Построим график зависимости скорости распада от времени. Получим кривую линию. Как убедиться в том,

[1] В физических справочниках часто вместо среднего времени жизни $\bar{t}$ указывают период полураспада $T\approx 0{,}69\ \bar{t}$ (см. § 1).

что эта кривая представляет собой график показательной функции? Для этого подсчитаем логарифмы полученных значений скорости распада и по этим данным построим график зависимости величины $\ln n(t)$ от времени t. В результате должна получиться прямая линия, а это нетрудно проверить на глаз. Многочисленные эксперименты действительно дают прямую.

Таким образом, $\ln n(t)$ есть линейная функция времени, т. е.

$$\ln n(t) = a + bt. \tag{8.2.1}$$

А это значит, что $n(t) = e^{a+bt} = e^a e^{bt} = ce^{bt}$. На графике величина b получается отрицательной: $b = -\omega$, где ω — вероятность распада. Следовательно, опыт подтверждает основной результат предыдущего параграфа и дает возможность определить ω, подсчитав тангенс угла наклона прямой (1) к оси t.

Надо подчеркнуть, что этот результат, по существу, весьма удивителен [2]. Представим себе N_0 радиоактивных атомов, приготовленных одновременно в начальный момент $t=0$. Все они приготовлены одинаковым способом и одновременно. Мы знаем, что радиоактивные атомы неустойчивы, способны распадаться. Можно допустить, что распад атомов требует определенного времени. Представим себе, что после приготовления атомы должны как-то созреть для распада. Но в таком случае мы должны были бы ожидать, что все атомы будут, не распадаясь, созревать в течение определенного времени, а по истечении этого времени, созрев, одновременно распадутся. Представим себе какие-то модели пушек с натянутыми пружинами и зубчатыми колесами (или часовыми механизмами), способные при определенном положении зубчатых колес (или часовых механизмов) силой пружины выбрасывать снаряд. Выстреливание снаряда можно назвать «распадом» модели. При одинаковом устройстве всех моделей модели, изготовленные в одно и то же время, выбрасывают снаряд (т. е. распадаются) по истечении одинакового срока.

Такая картина распада модели (вылета снаряда) не имеет ничего общего с поведением радиоактивного атома. Изготовленные в одно и то же время атомы распадаются в самые разнообразные моменты. Подсчитаем, например, сколько процентов распадается за время, не превосходящее среднего времени жизни. Из (1.2) находим, что скорость распада (количество атомов, распавшихся в единицу времени) есть $\frac{dN}{dt} = -\omega N_0 e^{-\omega t}$ (ясно, что $dN<0$). Изменение числа атомов за время dt равно

$$\frac{dN}{dt}dt = dN = -\omega N_0 e^{-\omega t} dt,$$

а за время от $t=0$ до $t=\bar{t}$ распадается

$$M = \int_0^{\bar{t}} \omega N_0 e^{-\omega t} dt = -N_0 e^{-\omega t}\Big|_0^{\bar{t}} = N_0 (1 - e^{-\omega \bar{t}}) \text{ атомов.}$$

Так как $\omega = \frac{1}{\bar{t}}$, то

$$M = N_0\left(1 - \frac{1}{e}\right) \approx 0{,}63 N_0.$$

Значит, за время от начального момента $t=0$ до момента $\bar{t}$ распадается 63% атомов. Аналогично подсчитываем, что за время от $\bar{t}$ до $2\bar{t}$ распадается 23% атомов, а за остальное время, после момента $2\bar{t}$, — всего 14% атомов.

На рис. 1 изображены зависимости от времени числа распадов в единицу времени для радиоактивных атомов (кривая *1*) и для моделей пушек, о которых шла речь выше (кривая *2*). При этом кривая *2* имеет определенную ширину. Можно представить себе, что модели изготовлены не совсем точно и поэтому стреляют не совсем одновременно. Чем точнее выполнены модели, тем у́же кривая *2* на рис. 1.

Площадь под кривой представляет собой общее число всех распавшихся атомов для одной кривой и общее число всех моделей для второй кривой. Можно взять число моделей, равное числу атомов. Тогда обе кривые будут ограничивать одинаковую площадь.

---

[2] Нильс Бор в докладе о радиоактивных превращениях говорил в этой связи: «Смысл дискуссий о средней продолжительности жизни атомов без указания на определенный момент времени состоит в том, что они, так сказать, не стареют до тех пор, пока не начнут распадаться; следовательно, одинаковая возможность распада существует в любой момент их жизни».

Абсциссы центра тяжести обеих кривых также одинаковы [3]; это значит, что наши модели пушек таковы, что их среднее время жизни (до выстрела) совпадает со средним временем жизни рассматриваемых радиоактивных атомов.

Таким образом, мы сделали все, что было в наших силах, чтобы добиться сходства кривых: взяли столько моделей и с таким механизмом, чтобы общее число моделей и атомов и среднее время жизни моделей и атомов были одинаковы. И тем не менее получились кривые, необычайно резко отличающиеся по форме! Эксперимент с радиоактивными ядрами с абсолютной неопровержимостью отбрасывает тот тип кривой, который получается для моделей пушек. Чем точнее ставятся опыты, тем с большей точностью подтверждается именно закон (1.2).

Сравнение с моделями нам понадобилось для того, чтобы не принимать как должную и естественную зависимость (1.2) для радиоактивного распада, чтобы вызвать чувство удивления и любопытства, вызвать вопрос: «А почему же, в самом деле, радиоактивный распад идет таким образом?».

Каков физический смысл величины вероятности распада? В далеком прошлом, в начале века, иногда высказывались предположения, что радиоактивный распад требует еще какого-то внешнего воздействия, например попадания извне какой-то частицы. В этом случае можно было бы себе представить, что один атом распался раньше, а другой позже в зависимости от того, в который из них раньше попала частица. Но эта гипотеза не соответствует фактам: радиоактивный распад идет с одинаковой скоростью в самых различных условиях, не зависит от температуры и частоты столкновений атомов между собой, не зависит от действия космических лучей; при радиоактивном распаде в точности сохраняется энергия, что также опровергает представление о каком-то внешнем воздействии при распаде.

Вторая возможная гипотеза заключается в предположении, что в действительности в начальный момент изготовления радиоактивных атомов они уже были не совсем одинаковы и именно поэтому распадаются в разное время. Это предположение соответствует картине моделей, которые выпускаются с часовыми механизмами, установленными на разное время. Подобная гипотеза предполагает, что точное знание состояния каждого атома полностью определяет всю его дальнейшую историю и, в частности, точно определяет, когда именно распадается данный атом. Если атомы распадаются через разное время после их приготовления, значит, так каждому из них на роду было написано: разные атомы одного и того же радиоактивного вещества были изготовлены уже неодинаковыми, что предопределило разное время их распада.

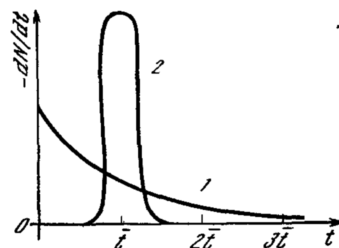


Рис. 8.2.1

Такая позиция также не выдерживает критики. С этой точки зрения при каждом конкретном способе получения атомов радиоактивного элемента должна получаться своя зависимость скорости распада от времени. Опыт опровергает это предположение.

Один и тот же вид радиоактивных атомов часто можно получить различными способами: например, атомы ^{99}Mo (молибден с атомной массой 99) получаются в атомных котлах при делении урана. Такие же атомы раньше были получены при действии ядер тяжелого водорода (дейтерия) на атомы обычного, встречающегося в природе, нерадиоактивного молибдена. В настоящее время известно множество примеров того, как один и тот же вид радиоактивных атомов получается разными способами. Опыт показывает, что независимо от способа получения атомов всегда скорость распада дается формулой (1.2) с постоянным значением ω, характеризующим данный вид атома. Следовательно, опыт доказывает, что во всех случаях справедливо наше основное уравнение $\frac{dN}{dt} = -\omega N$, где характеристика ω атомов данного

[3] Как будет показано в § 9.12, это следует из формулы (1.5).

вида от нас абсолютно не зависит. В уравнении (1.1) заключено очень большое содержание: все радиоактивные атомы в точности одинаковы. Вероятность распада не зависит от того, когда и как они были получены. Свежеполученные 100 атомов распадаются совершенно так же, как и в том случае, если приготовить 10^6 атомов, выждать время, нужное для того, чтобы уцелели 100 атомов, и рассмотреть эти 100 оставшихся атомов [4].

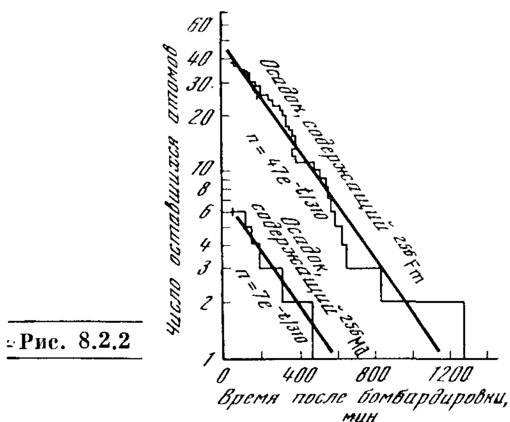


Рис. 8.2.2

Что же необычного в том, что 100 атомов с данной атомной массой и числом электронов всегда одинаковы? Если бы это были нерадиоактивные атомы, то и в самом деле удивляться было бы нечему. Но для радиоактивных атомов это удивительно. Вспомним, что из 100 атомов 63 распадаются за время $\bar{t}$, а остальные 37 — после $\bar{t}$, т. е. неожиданно, хотя все атомы и одинаковы, время распада у них различно.

Это удивление плодотворно. В явлении радиоактивного распада уже проявляются особенности законов поведения атомов и ядерных частиц, отличающихся от законов движения тел, с которыми мы встречаемся в классической механике и в повседневной жизни. Эти особенности изучаются в квантовой механике, которой мы в этой книге коснуться не можем. Нашей целью является более скромная задача: показать, что постановка вопроса о необходимости выработки новых представлений, отличающихся от обычной механики, следует из очень простых, известных каждому школьнику фактов о радиоактивности. Чтобы понять недостаточность старых представлений, надо только суметь задуматься, суметь удивиться.

В автобиографии Альберт Эйнштейн — величайший физик XX века — подчеркивает изумление, ощущение чуда, которое он испытал, когда впервые увидел компас, увидел таинственное действие магнитной силы, проникающей через бумагу, дерево, землю и воздействующей на стрелку компаса без прямого соприкосновения с ней. Он пишет о том, что такое удивление является сильнейшим побудительным мотивом для исследования. Он пишет о любознательности, «которую современные методы обучения почти совсем удушили». Сам Эйнштейн показал исключительную способность удивляться и черпать вдохновение и побуждение к созданию теорий из самых общеизвестных фактов. Так, в основе гениальной общей теории относительности лежит удивление перед фактом падения различных тел с одинаковым ускорением.

Понятно, что одной постановки вопроса недостаточно, и Эйнштейн соединял способность ставить вопрос с умением решить его, с владением всей нужной математической техникой. И все же именно умение удивиться и поставить новый вопрос там, где многие другие не видели ничего замечательного, сделало Эйнштейна наиболее выдающимся физиком нашего столетия.

Может быть, приведенный выше разбор радиоактивного распада послужит читателю примером того, какие глубины можно увидеть за простыми фактами и формулами.

В заключение параграфа приведем в качестве примера кривые радиоактивного распада (рис. 2), полученные экспериментально в 1955 г. Г. Сиборгом и его сотрудниками (США), которые впервые наблюдали 101-й элемент периодической системы, названный ав-

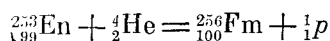
---

[4] Характерной особенностью показательной функции является подобие части ее графика всей кривой. Действительно, начнем новый отсчет времени с момента t_1. Время, отсчитанное с этого момента, обозначим τ, т. е. $\tau = t - t_1$, $t = t_1 + \tau$. Тогда $N = N_0 e^{-\omega t} = N_0 e^{-\omega (t_1 + \tau)} = N_0 e^{-\omega t_1} e^{-\omega\tau} = N_1 e^{-\omega\tau}$, где $N_1 = N_0 e^{-\omega t_1}$ — число атомов в момент t_1. Таким образом, закон распада оставшихся от предыдущего распада частиц (их число равно $N_1 = N_0 e^{-\omega t_1}$) совершенно такой же, как и закон распада свежеполученных N_1 частиц.

торами менделевием (химический символ Md) в честь великого русского химика Д. И. Менделеева.

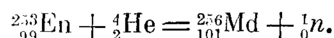
Сиборг со своими сотрудниками облучали нейтронами 98-й элемент, к а л и ф о р н и й, с атомной массой 252; нейтрон захватывался ядром атома и образовывался калифорний-253. Калифорний-253 испускал электрон и превращался в 99-й элемент — э й н ш т е й н и й (En) с той же атомной массой 253.

Около 10^9 атомов эйнштейния (т. е. $4 \cdot 10^{-13}$ г) осаждали на золотую пластинку и подвергали в циклотроне бомбардировке α-частицами, т. е. ядрами гелия. При этом образовывался элемент с порядковым номером 100 — ф е р м и й (Fm) согласно ядерной реакции



$$^{253}_{99}\mathrm{En} + {}^{4}_{2}\mathrm{He} = {}^{256}_{100}\mathrm{Fm} + {}^{1}_{1}p$$

и м е н д е л е в и й согласно реакции



$$^{253}_{99}\mathrm{En} + {}^{4}_{2}\mathrm{He} = {}^{256}_{101}\mathrm{Md} + {}^{1}_{0}n.$$

В этой записи у каждого химического символа внизу слева поставлен номер в периодической системе Менделеева, т. е. число протонов в ядре. Число слева сверху указывает атомную массу, округленную до целого числа, т. е. сумму числа нейтронов и протонов в ядре. ${}^{4}_{2}\mathrm{He}$ — это ядро г е л и я, т. е. α-частица, ${}^{1}_{1}p$ — ядро атома водорода (протон), ${}^{1}_{0}n$ — нейтрон. При ядерной реакции сумма нижних чисел в левой и правой частях уравнения реакции одинакова, а также одинакова и сумма верхних чисел, так как при ядерных реакциях происходит только обмен нейтронами и протонами между ядрами.

После бомбардировки α- частицами золотая пластинка вместе с образовавшимися фермием и менделевием растворялась в кислоте, и фермий и менделевий выделялись химически. Именно периодический закон Менделеева, как пишет Сиборг, позволяет предвидеть заранее химические свойства элемента, никогда ранее в природе не существовавшего, никогда ранее не исследованного. После химического разделения производились измерения радиоактивного распада. Фермий с атомной массой 256 распадается с периодом полураспада около 3,5 ч. Он распадается на два ядра-осколка примерно равной массы, т. е. самопроизвольно делится (о делении см. § 3).

Верхняя кривая на рис. 2 показывает зависимость числа ядер фермия от времени опыта. По оси абсцисс отложено время в минутах, по оси ординат — число атомов, имеющихся в данный момент [5]. Ось ординат имеет неравномерную шкалу: расстояние от оси абсцисс пропорционально л о г а р и ф м у числа атомов. В частности, ось абсцисс ($y=0$) соответствует одному уцелевшему атому ($\ln 1 = 0$), а числу атомов, равному нулю, соответствует $-\infty$ на оси ординат. Распад каждого отдельного атома меняет число атомов на 1; в промежутке между двумя распадами число атомов постоянно. Поэтому при такой экспериментальной технике, когда регистрируется каждый отдельный распад, вместо плавной кривой получается ступенчатая ломаная линия, на которой каждому распаду соответствует вертикальная линия, соединяющая две ступени. Верхняя прямая линия на рис. 2 соответствует закону распада

$$n = n_0 e^{-t/\tau},$$

где $\tau = \frac{T}{\ln 2} \approx 5$ ч, $T \approx 3{,}5$ ч. Как видно из рис. 2, всего в опыте зарегистрировано 40 распадов фермия. Чем больше атомов, тем ближе ломаная к прямой; когда же остается меньше пяти атомов, то естественно, что вероятностный характер радиоактивного распада приводит к значительным отклонениям от показательного закона, справедливого для б о л ь ш о г о числа атомов.

Ядра менделевия после химического разделения быстро (за полчаса) захватывают атомный электрон и превращаются при этом в ядра фермия. Поэтому тот осадок, в котором оказался менделевий, при измерении его радиоактивности также дает распад атомов на два осколка с периодом полураспада 3,5 ч. Кривая распада фермия, получившегося из менделевия, рас-

[5] Число атомов, имеющихся в данный момент, непосредственно в этот момент подсчитать не удается. На опыте регистрируются распады атомов. Число атомов N в момент t подсчитывается по окончании опыта, когда все N атомов уже распались.

положена в левом нижнем углу рис. 2. В опыте наблюдено шесть распадов. Специальными опытами доказано, что эти шесть атомов не могли попасть в изменяемый осадок в качестве загрязнения фермием, а именно образовались из менделевия.

Всего в нескольких опытах Сиборг и его сотрудники наблюдали 17 атомов менделевия.

Приведенный выше пример не очень хорош для иллюстрации того, насколько точно выполняется при радиоактивном распаде экспоненциальный закон (1.2) (ср. также (1)). Опыты, доказывающие этот закон, успешно ставились с более распространенными радиоактивными веществами. Зато пример менделевия и фермия показывает замечательное экспериментальное искусство современных физиков, синтезирующих новые элементы и регистрирующих распад каждого отдельного атома.

В опытах Сиборга счетчик распадов менделевия был включен через усилитель в радиотрансляционную сеть института, и при каждом распаде все сотрудники института, работающие в разных лабораториях, на разных этажах, оповещались об успехе — о рождении (точнее, о зарегистрированной смерти) каждого атома нового элемента, созданного человеком (впрочем, еще раньше, чем была закончена работа, вмешалась пожарная охрана, и оповещение было прекращено).

В нашей стране синтезом и исследованием самых тяжелых элементов успешно занимается академик Г. Н. Флеров и его сотрудники в Объединенном институте ядерных исследований в г. Дубне под Москвой.

§ 3. Последовательный распад (радиоактивное семейство)

В ряде случаев радиоактивный распад приводит к образованию атомов, которые также радиоактивны, так что осуществляется цепочка распадов: атом вещества А превращается в атом вещества В, атом вещества В в свою очередь превращается в атом вещества С и т. д. Рассмотрим математическую задачу об определении зависимости от времени количеств веществ А, В, С и способы ее решения. Количество атомов веществ А, В, С, не распавшихся к моменту времени t, будем обозначать теми же буквами A, B, C; таким образом, A и т. д. у нас — функции переменной t, т. е. $A=A(t)$ и т. д.

Пусть *вероятности распада* веществ А, В, С равны соответственно ω, v, u. Тогда

$$\frac{dA}{dt}=-\omega A \tag{8.3.1}$$

(А называют *материнским* веществом). Напишем уравнение для вещества В (*дочернее* вещество). За единицу времени распадается vB атомов вещества В. Кроме того, за то же время происходит ωA распадов вещества А; а так как при каждом распаде атома вещества А образуется атом вещества В, то за единицу времени образуется ωA атомов вещества В. Поэтому

$$\frac{dB}{dt}=-vB+\omega A. \tag{8.3.2}$$

Аналогичные рассуждения дают

$$\frac{dC}{dt}=-uC+vB. \tag{8.3.3}$$

Уравнения (1)—(3) образуют *систему дифференциальных уравнений*. В данном случае эти уравнения можно решать одно за другим, имея каждый раз дело только с одним уравнением и одним неизвестным. Действительно, в уравнение (1) B и C не входят. Поэтому из него определяем $A(t)=A_0e^{-\omega t}$; здесь A_0 — количество атомов вещества А в начальный момент $t=0$. (Мы полагаем, что $B=C=0$ при $t=0$.)

Подставляя выражение для $A(t)$ в уравнение (2), получим уравнение, содержащее только одну неизвестную функцию B:

$$\frac{dB}{dt}=-vB+\omega A(t). \tag{8.3.4}$$

Как решить такое уравнение? Его решение можно найти, если рассмотреть сперва судьбу группы атомов вещества В, образовавшихся в один и тот же промежуток времени — от τ до $\tau+\Delta\tau$. Будем рассматривать число атомов этой группы ΔB, оставшихся «в живых», т. е. не распавшихся к моменту t, в зависимости от времени t. Для того чтобы различать между собой момент t, когда мы измеряем число атомов, и момент образования группы,

обозначаем эти моменты разными буквами: t и τ соответственно. В момент τ скорость образования атомов вещества В была $\omega A(\tau)$. За малый промежуток времени $\Delta\tau$ образовалось $\Delta B_0 = \omega A(\tau)\Delta\tau$ атомов вещества В.

Как зависит число атомов в интересующей нас группе от времени t? При $t < \tau$ оно равно нулю: интересующие нас атомы еще не образовались, так как не образовалась еще сама группа, $\Delta B=0$. Пусть $t > \tau$. Заметим, что с момента образования группы прошло время $t - \tau$. Вероятность распада вещества В есть v. Поэтому по истечении времени $t - \tau$ с момента образования группы число нераспавшихся атомов будет равно

$$\Delta B(\tau) = \Delta B_0 e^{-v(t-\tau)} = \omega A(\tau) e^{-v(t-\tau)} \Delta\tau.$$

Для того чтобы найти полное число атомов вещества В в момент времени t, надо сложить число атомов во всех группах, образовавшихся до момента t. Если брать $\Delta\tau$ (а значит, и ΔB) весьма малым, то сумма превратится в интеграл:

$$B(t) = \int_0^t \frac{\Delta B(\tau)}{\Delta\tau} d\tau = \int_0^t \omega A(\tau) e^{-v(t-\tau)} d\tau.$$

Отметим, что переменная интегрирования обозначена здесь через τ. Аргумент t, от которого зависит B, входит в интеграл в двух местах: как верхний предел интеграла и в выражение подынтегральной функции. При интегрировании по τ величину t следует рассматривать как постоянную. Поэтому можно записать

$$e^{-v(t-\tau)} = e^{-vt} e^{v\tau}$$

и вынести ωe^{-vt} из-под интеграла как множитель, не зависящий от τ. Сделав это, получим

$$B(t) = \omega e^{-vt} \int_0^t A(\tau) e^{v\tau} d\tau. \qquad (8.3.5)$$

Легко проверить, не вычисляя интеграла, что решение (5) удовлетворяет исходному уравнению (4) при любой зависимости $A(\tau)$. Действительно, найдем производную $\frac{dB(t)}{dt}$. По правилу дифференцирования произведения получим

$$\frac{dB(t)}{dt} = -\omega v e^{-vt} \int_0^t A(\tau) e^{v\tau} d\tau + \omega e^{-vt} \frac{d}{dt}\left(\int_0^t A(\tau) e^{v\tau} d\tau\right).$$

Так как по свойству производной от интеграла (см. § 3.3)

$$\frac{d}{dt}\left(\int_0^t A(\tau) e^{v\tau} d\tau\right) = A(t) e^{vt},$$

то

$$\frac{dB}{dt} = -\omega v e^{-vt} \int_0^t A(\tau) e^{v\tau} d\tau + \omega A(t) = -vB + \omega A.$$

Если же мы положим $A(\tau) = A_0 e^{-\omega\tau}$, то получим конкретное решение:

$$B(t) = \frac{A_0\omega}{v-\omega}\left(e^{-\omega t} - e^{-vt}\right). \qquad (8.3.6)$$

Решение можно было найти и не прибегая к рассмотрению отдельных групп атомов. Теперь, когда решение найдено, уже легко угадать математический прием, ведущий к цели. Решение (5) имеет вид

$$B(t) = e^{-vt} I(t), \qquad (8.3.7)$$

где $I(t)$ обозначает интеграл, зависящий от t. Итак, будем искать решение в виде произведения e^{-vt} на неизвестную функцию I и составим уравнение для I:

$$\frac{dB}{dt} = \frac{d}{dt}\left(e^{-vt} I\right) = -v e^{-vt} I + e^{-vt} \frac{dI}{dt}. \qquad (8.3.8)$$

Подставляя выражения (8) и (7) в уравнение (4), получим

$$e^{-vt} \frac{dI}{dt} = \omega A(t),$$

или

$$\frac{dI}{dt} = \omega e^{vt} A(t). \qquad (8.3.9)$$

По условию, в начальный момент $t=0$ имеем $B=0$, а значит, и $I=0$.

Решение уравнения (9) с этим начальным условием имеет вид

$$I(t)=\int_0^t \omega e^{v\tau}A(\tau)\,d\tau.$$

Таким образом, окончательно получаем

$$B(t)=e^{-vt}I(t)=e^{-vt}\int_0^t \omega A(\tau)\,e^{v\tau}d\tau. \quad (8.3.10)$$

В этой формуле существенно во избежание ошибок сохранять более строгие обозначения и не обозначать переменную интегрирования τ той же буквой, которой обозначен предел интегрирования t.

§ 4. Исследование решения для радиоактивного семейства

В предыдущем параграфе мы довели до конца решение задачи в случае **двух** радиоактивных веществ. Исследуем это решение для двух частных случаев:

1) короткоживущее материнское вещество А, долгоживущее дочернее вещество В;

2) долгоживущее материнское вещество А, короткоживущее дочернее вещество В.

Ниже мы будем наряду с вероятностями распада ω и v пользоваться средними временами жизни $\bar{t}_A=\frac{1}{\omega}$, $\bar{t}_B=\frac{1}{v}$. В первом случае, когда $\bar{t}_A \ll \bar{t}_B$, характер решения легко усмотреть без всяких расчетов и точных формул. Здесь весь процесс распадается на две стадии. Сперва, при t порядка $\bar{t}_A$ (а так как по условию $\bar{t}_A \ll \bar{t}_B$, то в первой стадии $t \ll \bar{t}_B$), происходит превращение А в В; вещество В за это время практически не распадается. В этой стадии количество В равно разности количеств А — начального и уцелевшего к моменту t:

$$B(t)=A_0-A(t)=A_0-A_0e^{-\omega t}=$$
$$=A_0(1-e^{-\omega t}),\quad t\ll \bar{t}_B.$$

В конце этого периода практически **все** вещество А превратилось в **В**, количество В становится равным начальному количеству A_0 материнского вещества, количество А обращается в нуль. После этого медленно, длительно происходит распад В:

$$B(t)=A_0e^{-vt},\quad t\gg \bar{t}_A.$$

Покажем, как эти качественные соображения вытекают из точной формулы. Для случая двух радиоактивных веществ А и В мы в предыдущем параграфе получили формулу

$$B(t)=A_0\frac{\omega}{v-\omega}\left(e^{-\omega t}-e^{-vt}\right).$$

В нашем случае $\bar{t}_A \ll \bar{t}_B$, $\omega \gg v$, поэтому удобнее переменить знаки, чтобы иметь дело с положительными величинами в скобках и в знаменателе дроби. Тогда получим

$$B(t)=A_0\frac{\omega}{\omega-v}\left(e^{-vt}-e^{-\omega t}\right). \quad (8.4.1)$$

Так как $v\ll\omega$, то $\frac{\omega}{\omega-v}\approx\frac{\omega}{\omega}=1$.

Выражение $e^{-vt}-e^{-\omega t}$ рассмотрим для двух последовательных стадий. Сначала, при $t\ll \bar{t}_B=\frac{1}{v}$, будет $vt\ll 1$. При этом $e^{-vt}\approx 1$. Так как t может быть величиной порядка $\bar{t}_A$, а ωt, следовательно, порядка единицы, то величину $e^{-\omega t}$ надо вычислять точно. Из формулы (1) следует

$$B(t)\approx A_0(1-e^{-\omega t}). \quad (8.4.2)$$

Во второй стадии, при $t\gg \bar{t}_A=\frac{1}{\omega}$, будет $\omega t\gg 1$. В этой стадии можно пренебречь малой величиной $e^{-\omega t}$: ведь $e^{-\omega t}$ мало не только по сравнению с единицей, но и по сравнению с e^{-vt}, так как $v\ll\omega$. Мы получим

$$B(t)=A_0e^{-vt}. \quad (8.4.3)$$

Таким образом, действительно, точная формула дает те результаты, которые были получены из простых качественных соображений.

Обратимся к случаю долгоживущего материнского вещества А и короткоживущего дочернего вещества В:

$$\bar{t}_A\gg \bar{t}_B,\quad \omega\ll v.$$

Рассмотрим стадию, когда со времени начала процесса прошло время t, значительно превышающее $\bar{t}_B$. В таком случае то вещество В, которое образовалось в начале процесса, к моменту

t уже полностью распалось. Так как В распадается быстро за малое время, то в каждый момент в наличии находится лишь вещество В, образовавшееся недавно. В рассматриваемом случае имеет место *установившееся*, или, как его иначе называют, *стационарное состояние*: вещество В образуется из А и тут же распадается, вещество В при этом не накапливается (потому что оно быстро распадается), но и не исчезает совсем (потому что из вещества А все время получается новое вещество В). Установившееся состояние характеризуется тем, что в единицу времени распадается примерно столько же атомов В, сколько новых атомов В образуется из А, так что здесь B мало меняется. Математически это условие совпадения числа распадающихся и возникающих атомов В записывается так:

$$vB \approx \omega A.$$

Отсюда

$$B(t) \approx \frac{\omega}{v} A(t) = \frac{\bar{t}_B}{\bar{t}_A} A(t). \qquad (8.4.4)$$

В установившемся состоянии мгновенное количество В пропорционально количеству А и составляет всегда одну и ту же малую долю А. Эта доля мала потому, что в рассматриваемом случае $\bar{t}_B \ll \bar{t}_A$ и, значит, $\frac{\bar{t}_B}{\bar{t}_A} \ll 1$, а иначе не было бы и самого установившегося состояния.

Как получить уравнение для установившегося состояния из точного дифференциального уравнения $\frac{dB}{dt} = -vB + \omega A$? Очевидно, если считать, что $\frac{dB}{dt}$ мало по сравнению с каждым из двух членов, стоящих в правой части, то приближенно, заменяя $\frac{dB}{dt}$ на 0, как раз и получим

$$0 = -vB + \omega A, \quad \text{т. е.} \quad vB = \omega A.$$

Рассмотрим теперь начало процесса. При $t=0$ имеем $A=A_0$, $B=0$. Значит, в начальный момент мы имеем дело не с установившимся состоянием, так как по формулам установившегося состояния вначале было бы

$$B_{ст} = \frac{\omega}{v} A_0$$

(значок «ст» у буквы B означает установившееся, стационарное состояние). В момент $t=0$ вещество В образуется со скоростью $\frac{dB}{dt} = \omega A_0$, а распада В в начальный момент вовсе не происходит, так как $B=0$.

Можно определить, за какое время t_1, при постоянной скорости нарастания вещества В, равной начальной скорости, будет достигнуто количество $B_{ст}$. Действительно, если скорость образования вещества остается постоянной, равной $\left(\frac{dB}{dt}\right)_{t=0}$, то

$$B = t\left(\frac{dB}{dt}\right)_{t=0}.$$

Полагая здесь $B_{ст} = \frac{\omega}{v} A_0$, $\left(\frac{dB}{dt}\right)_{t=0} = \omega A_0$, получим искомое время:

$$t_1 = \frac{\omega A_0}{v\omega A_0} = \frac{1}{v} = \bar{t}_B.$$

Таким образом, установившееся состояние достигается за время, приблизительно равное среднему времени распада вещества В (напомним, что наше предположение о постоянстве скорости образования вещества В выполняется лишь приблизительно). Из условия $\bar{t}_A \ll \bar{t}_B$ видно, что количество вещества А за это время изменится мало.

В целом приближенное рассмотрение в случае короткоживущего дочернего вещества дает следующие результаты:

$$B(t) = \left(\frac{dB}{dt}\right)_{t=0} t = \omega A_0 t \quad \text{при } t < \bar{t}_B,$$

$$B(t) = B_{ст} = \frac{\omega}{v} A_0 e^{-\omega t} = \omega \bar{t}_B A_0 e^{-\omega t} \quad \text{при } t > \bar{t}_B. \qquad (8.4.5)$$

Мы получаем зависимость $B=B(t)$ в виде двух линий: сначала прямой *1* (т. е. линейной функции), а затем экспоненты (графика показательной функции; см. рис. 1, составленный в предположении, что $\bar{t}_A = 10\, \bar{t}_B$; вместо экспоненциальной кривой здесь изображена близкая к ней прямая *2*). Нетрудно проверить, что при $t=\bar{t}_B$ две формулы (5) дают близкие значения.

Посмотрим, что дает точное решение (1) уравнения (3.4) в рассматриваемом случае, когда $v \gg \omega$, $\bar{t}_B \ll \bar{t}_A$. Пренебрежем ω по сравнению с v в зна-

менателе. При $\nu t \gg 1$ пренебрежем также и $e^{-\nu t}$ в скобках. Мы получим

$$B \approx A_0 \frac{\omega}{\nu} e^{-\omega t}, \qquad (8.4.6)$$

т. е. как раз установившееся решение.

Приближение решения к установившемуся определяем тем, насколько быстро убывает $e^{-\nu t}$. При совсем малом t (когда $\nu t < 1$, так что ωt и подавно мало), разлагая в ряд $e^{-\nu t}$ и $e^{-\omega t}$ и ограничиваясь двумя первыми членами, получим

$$B \approx A_0 \frac{\omega}{\nu - \omega}(1 - \omega t - 1 + \nu t) = A_0 \omega t, \qquad (8.4.7)$$

что тоже совпадает с приближенным результатом. Однако в действительности точная формула дает на графике одну плавную кривую *3* без изломов и разрывов (она изображена пунктиром на рис. 1). Приближение этой кривой к установившемуся решению зависит от того, насколько быстро уменьшается $e^{-\nu t}$. Так, для того чтобы $e^{-\nu t}$ давало поправку порядка 10%, нужно, чтобы было $\nu t \approx 2{,}3$, $t \approx 2{,}3 \frac{1}{\nu} = 2{,}3\bar{t}_B$. При этом из-за малости ωt считаем, что $e^{-\omega t} \approx 1$. Таким образом, действительно, переход от стадии начального нарастания к стадии, когда решение с достаточной точностью совпадает с установившимся, происходит за время, по порядку величины совпадающее с временем распада $\bar{t}_B$.

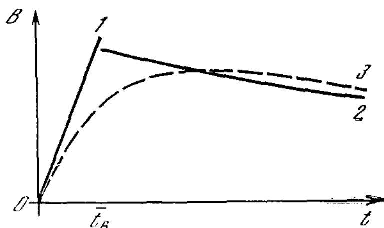


Рис. 8.4.1

Пример радиоактивного семейства очень поучителен в том отношении, что получение общего точного решения никоим образом нельзя считать концом работы над задачей. Построение приближенных теорий для различных предельных случаев есть совершенно необходимая часть всего исследования, и наличие точной формулы вовсе не заменяет приближенной теории. Приближенные, но ясные и наглядные представления служат для проверки точной формулы.

Приближенные теории дают нам такие важные новые качественные понятия, как понятие установившегося состояния. Такие понятия и лучше запоминаются, и обладают более широкой, чем точные формулы, областью применимости. Так, например, в случае радиоактивного семейства, состоящего из н е с к о л ь к и х поколений A → B → C → D, точная формула весьма громоздка. Однако если $\bar{t}_A$ больше всех других времен жизни $\bar{t}_B$, $\bar{t}_C$ и $\bar{t}_D$, то результаты, относящиеся к установившемуся состоянию, получаются так же просто, как и в случае д в у х веществ A и B.

Иногда наиболее легкий путь состоит в том, чтобы получить точное решение, справедливое, скажем в нашем примере, для л ю б ы х ν и ω, а затем, положив $\nu \ll \omega$ или $\nu \gg \omega$, вывести из этого общего решения путем математических преобразований результата более простые приближенные формулы для двух крайних случаев. Но на этом нельзя успокаиваться! Если простая приближенная формула получилась хотя и легким, но длинным путем, использующим общее решение, то должен существовать и другой, простой способ получения приближенной формулы. Нужно обязательно тренироваться в нахождении простых методов, потому что будут встречаться задачи, в которых нахождение точного общего решения непреодолимо сложно, и только простой приближенный метод позволит продвинуть вперед решение вопроса.

В практической работе точные формулы встречаются так же редко, как алгебраические, тригонометрические или любые иные уравнения, решения которых выражаются целыми числами, —и это несмотря на то, что в учебниках математики и физики большинство задач приводят к точным формулам, подобно тому как в задачниках для младших классов школы фигурируют лишь уравнения, которые решаются в целых числах.

Отметим, что представления о радиоактивных семействах объясняют странный результат, получившийся в упр.

1.5, относительно количества радия в прошлом: радий является потомком (правда, не прямым, а через несколько промежуточных веществ) урана-238. Поэтому нельзя считать имеющееся в настоящее время количество радия результатом распада первоначального радия. В действительности радий находится в установившемся состоянии с ураном. Из уравнения

$$B = \frac{\omega}{v} A$$

найдем, что содержание радия $B = 10^{-12}$ соответствует содержанию урана

$$A = \frac{t_A}{t_B} B = 3 \cdot 10^6 B = 3 \cdot 10^{-6}.$$

Мы нашли приближенно современное количество урана-238 в горных породах. Начальное содержание его $5 \cdot 10^9$ лет назад было вдвое больше, порядка $6 \cdot 10^{-6}$. Такие величины вполне правдоподобны в отличие от результатов упражнений к § 1.

§ 5. Цепная реакция деления урана

В 1938 г. Хан и Штрассман в Германии и супруги Жолио-Кюри во Франции показали, что при попадании нейтрона в ядро урана происходит деление ядра, при котором оно распадается на два больших осколка и, кроме того, испускает два-три новых нейтрона. Особенно активно делится уран с атомной массой 235. В природном уране содержится около 0,7% атомов урана-235 и 99,3% атомов урана-238 [1]. «Осколки» деления урана-235 представляют собой ядра средней атомной массы — от 75 до 160; заряд этих ядер лежит в пределах от 35 до 57, при этом сумма зарядов двух осколков всегда равна заряду ядра урана, т. е. 92 элементарным зарядам; сумма атомных масс двух осколков равна $235+1-\nu$, где 235 — атомная масса урана, 1 — атомная масса нейтрона, вызвавшего деление, ν — масса нейтронов, образовавшихся при делении. При делении выделяется большая энергия — порядка $6 \cdot 10^{10}$ Дж/г (на 1 г разделившегося урана). Сразу после деления осколки летят в противоположные стороны со скоростью около 10^9 см/с; затем они тормозятся и их кинетическая энергия превращается в тепло.

Источником этой энергии является электрическое отталкивание двух одноименно заряженных осколков. Пока ядро не разделилось на две части, ядерные силы притяжения частиц, из которых состоит ядро, уравновешивают электрическое отталкивание. Как только ядро разделилось на два отдельных осколка, отталкивание этих двух осколков уже ничем не уравновешивается и приводит к тому, что они разлетаются с большой скоростью. В плотном веществе осколки очень быстро останавливаются. Время их движения составляет от 10^{-13} до 10^{-12} с. При этом они проходят от 10^{-4} до 10^{-3} см. Нейтроны, образующиеся при делении, имеют скорости того же порядка, что и осколки (около $2 \cdot 10^9$ см/с).

Для практического использования энергии деления ядер решающее значение имеет тот факт, что при делении, вызванном **одним** нейтроном, получается **больше одного** нейтрона. Ясно, что если нейтроны не будут уходить из системы, то число их будет нарастать в геометрической прогрессии с течением времени, т. е. по закону показательной функции. По такому же закону пропорционально числу нейтронов будет нарастать и скорость выделения энергии. При этом если даже в начале процесса было мало нейтронов, то очень скоро число их возрастет настолько, что энергия будет выде-

[1] Вслед а результатами Хана—Штрассмана и Жолио-Кюри в 1939 г. в лаборатории И. В. Курчатова в Ленинграде советские ученые Г. Н. Флеров и К. А. Петржак показали, что уран-238 способен делиться *самопроизвольно*, без попадания в него нейтрона, хотя вероятность этого весьма мала. Вероятность радиоактивного распада (с испусканием α-частицы) урана-238, соответствующая периоду полураспада $4,5 \cdot 10^9$ лет, примерно равна $\omega = 5 \cdot 10^{-18}$ $с^{-1}$, а вероятность самопроизвольного деления урана-238 в 10^6 раз меньше, т. е. ее можно оценить величиной $5 \cdot 10^{-24}$ $с^{-1}$. Таким образом, в 1 кг урана, содержащем примерно $2,5 \cdot 10^{25}$ атомов, в 1 с происходит около 10^7 актов радиоактивного распада и всего 10 самопроизвольных делений. Однако в самых тяжелых элементах самопроизвольные деления становятся уже главным, наиболее вероятным процессом распада (см. конец § 2, в частности рис. 2.2, где приведена кривая распада менделевия). В рассматриваемом ниже вопросе о цепной реакции самопроизвольное деление роли не играет.

ляться со скоростью, удобной для практического использования (например, в качестве источника энергии атомных электростанций или в двигателях ледокола); спустя еще совсем небольшое время выделение энергии возрастет настолько, что произойдет атомный взрыв. В действительности часть нейтронов выходит из системы, а часть нейтронов может захватываться другими ядрами, не вызывая деления; пользуясь этим, можно регулировать число нейтронов, в частности можно добиться у с т а н о в и в ш е г о с я   с о с т о я н и я системы, при котором число образующихся в единицу времени нейтронов равно числу исчезающих нейтронов, так что число нейтронов, находящихся в системе, с течением времени не изменяется и выделение энергии происходит с постоянной скоростью. Для использования атомной энергии в мирных целях нужен именно такой режим работы.

Наша ближайшая задача — составить и исследовать уравнение, описывающее зависимость числа нейтронов в системе от времени, это и будет сделано в следующих параграфах.

§ 6. Размножение нейтронов в большой массе

Получим сперва уравнение для изменения числа нейтронов со временем в очень большой системе (например, в большом слитке урана-235), когда выход нейтронов наружу можно не учитывать [1]. Скорости всех нейтронов приближенно можно считать одинаковыми; обозначим эту скорость через v.

Деление ядра происходит примерно в половине всех случаев, когда нейтрон попадает в ядро урана-235. Во второй половине случаев нейтрон вылетает обратно, оставляя ядро в прежнем состоянии, и при этом число нейтронов не меняется. Ядро урана представляет собой шарик радиуса R порядка 10^{-12} см.

Как часто нейтрон, летящий внутри металла, попадает в ядро урана?

За малое время dt нейтрон проходит путь vdt. Представим себе цилиндр, осью которого является путь, пройденный нейтроном; радиус цилиндра равен радиусу ядра урана R. Нейтрон сталкивается с теми ядрами, центры которых находятся внутри цилиндра, при этом путь нейтрона удален от центра ядра на расстояние меньше R; поэтому этот нейтрон задевает ядро — попадает в него. Объем цилиндра равен $\pi R^2 vdt$.

На рис. 1 приведена схема, поясняющая эти соображения. Путь нейтрона показан пунктирной линией, проходящей по оси цилиндра. В случае $А$ центр ядра находится внутри цилиндра и нейтрон задевает ядро. В случае $Б$ центр ядра расположен вне цилиндра и нейтрон пролетает, не задевая ядро.

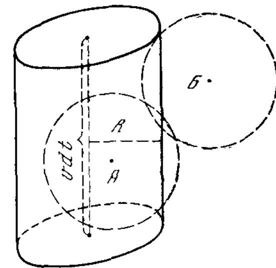


Рис. 8.6.1

В металлическом уране в единице объема содержатся N атомов и, следовательно, N ядер (размерность N есть 1/см^3). Поэтому в интересующем нас объеме $\pi R^2 vdt$ имеется $N\pi R^2 vdt$ ядер. Столько же будет попаданий нейтрона в ядро за малый промежуток времени dt.

Не всякое попадание нейтрона вызывает деление ядра. Пусть α — доля тех случаев попадания нейтрона в ядро, когда это попадание вызвало деление (в случае урана-235 $\alpha \approx 1/2$). Тогда число делений за время dt равно $N\alpha\pi R^2 vdt$.

Величина $\alpha\pi R^2$ имеет размерность площади, так как α и π безразмерны. Она называется *сечением* деления; мы будем обозначать ее символом $\sigma_д$.

Если внутри массы металлического урана находятся n нейтронов, то число делений за время dt равно

$$nN\sigma_д vdt. \tag{8.6.1}$$

При каждом делении образуется ν нейтронов, но за счет поглощения одного нейтрона. Таким образом, число нейтронов при каждом делении увеличивается на $\nu-1$. Числу делений (1) соответствует следующее изменение числа нейтронов:

$$dn = nN(\nu - 1)\sigma_д vdt. \tag{8.6.2}$$

[1] Мы будем рассматривать самый простой случай металлического урана-235 без графитового замедлителя и т. п.

Из (2) следует

$$\frac{dn}{dt} = nN(\nu - 1)\sigma_д v.$$

Положим

$$N(\nu - 1)\sigma_д v = a; \qquad (8.6.3)$$

тогда

$$\frac{dn}{dt} = an.$$

Как мы уже знаем, решение этого дифференциального уравнения имеет вид

$$n(t) = n_0 e^{at}, \qquad (8.6.4)$$

где n_0 — число нейтронов в системе при $t=0$.

Таким образом, если число нейтронов в системе меняется только вследствие деления, то это число растет в геометрической прогрессии, если время растет в арифметической прогрессии. Действительно, если взять несколько равноотстоящих моментов времени

$$t_1,\ t_1 + \Delta t,\ t_1 + 2\Delta t,\ t_1 + 3\Delta t, \ldots, \qquad (8.6.5)$$

то соответствующие числа нейтронов равны

$$n_1 = n_0 e^{at_1},\ fn_1,\ f^2 n_1,\ f^3 n_1, \ldots,$$
где $f = e^{a\Delta t}$. (8.6.5a)

Отметим, что такой способ описания интересующего нас процесса — рост числа нейтронов в геометрической прогрессии — часто встречается в популярной литературе (ср. сноску 3 на с. 125). Специалисты (физики и техники) пользуются подобными (5), (5а) записями редко; они чаще и охотнее говорят об *экспоненциальном* законе возрастания. Экспоненциальный рост (4) характеризуется *скоростью роста* a.

Какова размерность a? В формуле (4) at стоит в показателе степени. Поэтому at — безразмерная величина и, следовательно, размерность a есть $с^{-1}$. Этот же результат можно получить, припомнив, что

$$a = N(\nu - 1)\sigma_д v,$$

где N выражено в 1/см^3, $\sigma_д$ — в см^2, v — в см/с.

Найдем приближенное значение постоянной величины a. Плотность урана приблизительно равна 18 г/см^3. Число N ядер в 1 см^3 подсчитаем, используя число Авогардо: 1 грамм-атом любого вещества содержит $6 \cdot 10^{23}$ атомов, следовательно, 235 г урана-235 содержат $6 \cdot 10^{23}$ атомов, т. е. $6 \cdot 10^{23}$ ядер. В 1 см^3 содержится $\frac{18}{235} \cdot 6 \cdot 10^{23} \approx 4 \cdot 10^{22}$ ядер, т. е. $N \approx 4 \cdot 10^{22}$ 1/см^3. Подставим в выражение для a среднее значение $\nu \approx 2{,}5$, $v \approx 2 \cdot 10^9$ см/с, $\sigma_д = \frac{1}{2}\pi(10^{-12})^2 \approx 1{,}6 \cdot 10^{-24}$ см^2.

Получим

$$a \approx 4 \cdot 10^{22} \cdot 1{,}5 \cdot 1{,}6 \cdot 10^{-24} \cdot 2 \cdot 10^9 \approx 2 \cdot 10^8\ с^{-1},$$

$$\frac{1}{a} \approx 5 \cdot 10^{-9}\ с.$$

Таким образом, если нейтроны не вылетают из системы, то количество нейтронов возрастает в e раз примерно за $5 \cdot 10^{-9}$ с.

При такой скорости возрастания за 1 мкс, т. е. за 10^{-6} с, количество нейтронов возрастает в $e^{2 \cdot 10^8 \cdot 10^{-6}} = e^{200}$ раз, т. е. примерно в $10^{0{,}43 \cdot 200} = 10^{86}$ раз.

Одна тонна урана-235 содержит около $2{,}5 \cdot 10^{27}$ ядер. Поэтому если нейтроны не выходят из системы, то это количество урана разделится меньше чем за 1 мкс после попадания первого нейтрона. Такой процесс представляет собой взрыв огромной силы.

Для использования деления с целью получения (и использования) энергии такая скорость процесса недопустима. Необходимо путем искусственного выведения нейтронов из системы уменьшить скорость нарастания их числа.

§ 7. Вылет нейтронов

Представим себе массу урана-235 в виде шара радиуса r. Мы должны составить уравнение, описывающее изменение числа n нейтронов, находящихся внутри этого шара. Предположим для простоты, что шар закреплен на какой-то тонкой подставке, так что вокруг него полная пустота; нейтрон, покинувший шар, уже никогда не возвращается обратно.

Как определить поток нейтронов, т. е. число нейтронов, выходящих за пределы шара в единицу времени? Сделаем грубый подсчет. Рассмотрим малый проме-

жуток времени dt. За это время каждый нейтрон проходит путь длиной vdt. Где в этом шаре находятся те нейтроны, которые покинут его за время dt? Очевидно, что они должны находиться внутри шара в тонком слое, прилегающем к поверхности шара, но на расстоянии не более vdt от поверхности, иначе за время dt они не успеют дойти до поверхности, пересечь ее и выйти наружу. Но и те нейтроны, которые находятся внутри слоя толщиной vdt, тоже не все успеют выйти за пределы шара за время dt, так как не у всех нейтронов внутри слоя скорость направлена по радиусу наружу; при первоначальном, грубом подсчете мы не будем учитывать последнее обстоятельство.

Найдем число нейтронов в слое. Во всем шаре имеется n нейтронов. Объем шара $V={}^4/_3\pi r^3$; объем интересующего нас тонкого слоя у поверхности приближенно равен $Svdt$, если vdt мало. Здесь $S=4\pi r^2$ (поверхность шара).

Средняя *плотность* нейтронов, т. е. число нейтронов в единице объема, равна $C=\frac{n}{V}$. Предположим, что в тонком слое у поверхности плотность не отличается от средней. Тогда число нейтронов в этом слое равно

$$CSvdt=\frac{nS}{V}vdt.$$

Поэтому для *потока* (числа нейтронов, выходящих в единицу времени) мы получим

$$q=\frac{nS}{V}v=\frac{n\cdot 4\pi r^2}{{}^4/_3\pi r^3}v=\frac{3v}{r}n.$$

В действительности у поверхности плотность нейтронов меньше средней плотности; к тому же, как отмечалось выше, скорости нейтронов имеют различные направления и не все нейтроны выходят из шара. Поэтому в действительности поток нейтронов меньше, чем мы получили:

$$q=\frac{3kv}{r}n, \qquad (8.7.1)$$

где k — численный коэффициент, $k<1$. В § 10 путем сопоставления с опытом будет показано, что для шара из урана-235 в условиях, когда становится возможен самоподдерживающийся процесс, k близко к 0,3. Если внутри шара не происходит деления ядер и не рождаются новые нейтроны, то для числа нейтронов внутри шара получается уравнение $\frac{dn}{dt}=-q$, или в силу (1) уравнение

$$\frac{dn}{dt}=-\frac{3kv}{r}n.$$

Обозначая

$$\frac{3kv}{r}=b, \qquad (8.7.2)$$

получим

$$\frac{dn}{dt}=-bn.$$

Решение такого уравнения нам хорошо известно:

$$n=n_0e^{-bt}. \qquad (8.7.3)$$

Среднее время пребывания нейтронов внутри шара согласно (3) равно

$$\bar{t}=\frac{1}{b}=\frac{r}{3kv}. \qquad (8.7.4)$$

Если величина $k=0{,}3$, то (4) дает $\bar{t}\approx\frac{r}{v}$. Поэтому среднее время приблизительно равно времени, за которое нейтрон, движущийся со скоростью v, проходит путь, равный радиусу шара r.

Точное рассмотрение вылета нейтронов требует весьма трудоемких расчетов. Очень важно с первых шагов обучения привыкнуть приближенно определять все интересующие нас величины. Точный расчет часто бывает действительно очень громоздким и требует совсем другого объема знаний; иногда здесь необходим коллективный труд нескольких человек с использованием ЭВМ и т. д. Значит ли это, что учащийся, занимающийся самостоятельно, не может и помышлять о рассмотрении вопроса? Всегда есть простые, хотя и грубые (вроде приведенного выше), способы приближенного, «прикидочного» подхода к задаче. Не произвести приближенного расчета и ссылаться на то, что точный расчет труден, — значит просто прикрывать такой ссылкой свою нерешительность и робость. А чаще всего именно робость мешает начинающим (иногда даже — не только начинающим) ученым и изобретателям!

§ 8. Критическая масса

До сих пор мы рассматривали отдельно два процесса: размножение нейтронов без учета их вылета и вылет нейтронов без учета их размножения.

Рассмотрим теперь систему, в которой нейтроны и размножаются, и могут покидать ее. В единицу времени в системе, как мы знаем, образуются an нейтронов и bn нейтронов вылетают из нее. Так как изменение числа нейтронов в единицу времени есть $\frac{dn}{dt}$, то

$$\frac{dn}{dt}=an-bn,$$

или

$$\frac{dn}{dt}=cn, \tag{8.8.1}$$

где $c=a-b$. При данном начальном количестве нейтронов n_0 уравнение (1) имеет решение

$$n=n_0e^{ct}. \tag{8.8.2}$$

Это решение приводит к совершенно различным результатам при положительных и отрицательных значениях c. Действительно, из (2) видно, что при $c<0$ число нейтронов n тем меньше, чем больше t; в этом случае n стремится к нулю при неограниченном росте t. Если же $c>0$, то n растет с ростом t; в этом случае число n становится неограниченно большим, если только t достаточно велико. Только вмешательство новых физических факторов, не учтенных в уравнении (1), может приостановить рост n.

Таким образом, значение $c=0$ есть «критическое значение»: оно разделяет решения разного типа — с растущим и с уменьшающимся количеством нейтронов. Так как $c=a-b$, то при данном a можно говорить о критическом значении b: $b_{кр}=a$, так как при $b<b_{кр}$ величина $c=a-b>0$, а при $b>b_{кр}$ величина $c=a-b<0$. Величина a определяется свойствами делящегося вещества: согласно формуле (6.2) $a=Nv\sigma_д\times\times(\nu-1)$. Величина b зависит от количества взятого делящегося вещества:

$$b=\frac{3kv}{r}.$$

Поэтому вводят понятие *критического значения* радиуса $r_{кр}$, при котором $b=b_{кр}=a$. Из формул (6.2) и (7.2) следует, что $\frac{3kv}{r_{кр}}=Nv\sigma_д(\nu-1)$, откуда

$$r_{кр}=\frac{3k}{N\sigma_д(\nu-1)}.$$

Масса шара, радиус которого равен $r_{кр}$, называется *критической массой* и обозначается через $m_{кр}$. Ясно, что

$$m_{кр}=\frac{4}{3}\pi r_{кр}^3\rho, \tag{8.8.3}$$

где ρ — плотность делящегося вещества [1].

При $r>r_{кр}$ (или, что то же, $m>m_{кр}$) будет $c>0$, т. е. имеет место размножение нейтронов. При $r<r_{кр}$ (т. е. при $m<m_{кр}$) будет $c<0$ — первоначально взятое количество нейтронов (экспоненциально) уменьшается. Пусть взят шар радиуса r. Его поверхность S и объем V равны

$$S=4\pi r^2,\quad V=\frac{4}{3}\pi r^3$$

и, значит,

$$\frac{S}{V}=\frac{4\pi r^2}{\frac{4}{3}\pi r^3}=\frac{3}{r}.$$

Если r мало, то отношение $\frac{S}{V}$ велико; если же r велико, то это отношение мало. Неудивительно, что при малом радиусе, когда отношение поверхности к объему велико, вылет нейтронов усиливается, условия для размножения нейтронов хуже. Удивительна та резкость, с которой меняется количество нейтронов при изменении b: если $b>b_{кр}$, то через некоторое время количество нейтронов обращается практически в нуль независимо от того, имеем ли мы $b=1{,}01b_{кр}$ или $b=2b_{кр}$. Если $b<b_{кр}$, то число нейтронов неограниченно возрастает и при $b=0{,}99b_{кр}$, и при $b=0{,}5b_{кр}$, хоть и с разной скоростью. Именно поэтому говорят о «критическом» значении b, «критическом» значении r или «критическом» значении массы. Масса, большая критической, называется *надкритической*, а масса, меньшая критической,— *подкритической*.

---

[1] Мы по-прежнему считаем, что рассматривается шарообразная масса делящегося вещества (например, урана).

На рис. 1 представлены кривые $n=n_0 \exp[(a-b)t]$ при нескольких значениях b. Построим кривые зависимости n от b для нескольких выбранных значений времени t. При расчете принято

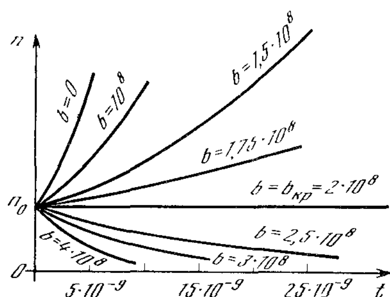


Рис. 8.8.1

$a=2\cdot 10^8$ 1/с. На рис. 2 показаны кривые $n=n(b)$, отвечающие значениям $t=5\cdot 10^{-9}$, $t=15\cdot 10^{-9}$ и $t=30\cdot 10^{-9}$ с.

Кривые, отвечающие значениям $t=15\cdot 10^{-9}$ и $t=30\cdot 10^{-9}$ с, пересекаются

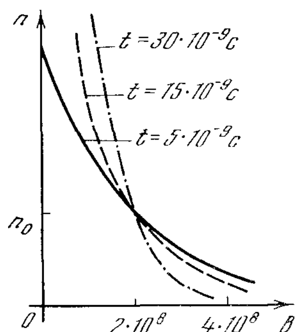


Рис. 8.8.2

с осью ординат (линией $b=0$) при значениях $n=20n_0$ и $n=400n_0$ соответственно.

Как видно из рис. 1 и 2, чем больше время t, тем сильнее расходятся кривые $n(t)$ (см. рис. 1), тем круче кривые $n(b)$ (см. рис. 2), тем резче проявляется критичность значения $b=2\cdot 10^8$ (в данном примере).

Если взять $t > 10^{-6}$ с, то кривую $n(b)$ нельзя будет отличить от вертикальной прямой $b=b_{кр}=2\cdot 10^8$; $n=0$ при $b > b_{кр}$, $n=\infty$ при $b < b_{кр}$.

§ 9. Подкритическая и надкритическая массы при непрерывном источнике нейтронов

В предыдущем параграфе была рассмотрена задача об изменении со временем числа нейтронов при данном начальном числе n_0 нейтронов. Поставим теперь несколько иную задачу. Пусть в начальный момент $t=0$ число нейтронов равно нулю и в этот момент включен *источник нейтронов*, испускающий q_0 нейтронов в единицу времени. Эта задача приводит к уравнению

$$\frac{dn}{dt}=cn+q_0, \tag{8.9.1}$$

где $c=a-b$. Мы ищем решение этого уравнения с начальным условием: $n=0$ при $t=0$.

Способ решения подобной задачи был изложен в § 3. Вкратце повторим ход рассуждений.

Ищем количество нейтронов в момент t. Весь промежуток времени от 0 до t разбиваем на отдельные промежутки $\Delta\tau$. Рассмотрим один такой промежуток времени от τ до $\tau+\Delta\tau$. Источник выдал за это время количество нейтронов, равное $q_0\Delta\tau$. Если бы источник действовал только в течение одного промежутка времени $\Delta\tau$, то мы имели бы дело с задачей предыдущего параграфа с начальным количеством нейтронов $n_0=q_0\Delta\tau$; единственное отличие состоит в том, что эти нейтроны выделились в момент $t=\tau$, а не в момент $t=0$. Поэтому вместо решения $n=n_0e^{ct}$ мы получили бы решение

$$n=n_0e^{c(t-\tau)}=q_0\Delta\tau e^{c(t-\tau)}$$

(это решение относится к моменту $t>\tau$; при $t<\tau$ имеем $n=0$), так как ясно, что изменение количества нейтронов зависит именно от времени, которое прошло после того момента, когда было задано их начальное число, т. е. в данном случае от величины $t-\tau$.

На самом деле источник действует непрерывно в течение всего времени от 0 до t. Значит, надо суммировать вклады всех нейтронов, испущенных источником в различные промежутки времени $\Delta\tau$, причем все эти промежутки вместе составляют весь интервал времени от 0 до t. Такая сумма при малых $\Delta\tau$ обращается в интеграл; поэтому

$$n(t)=\int_0^t q_0e^{c(t-\tau)}d\tau.$$

Этот интеграл легко вычислить:

$$n(t)=q_0e^{ct}\int_0^t e^{-c\tau}d\tau=q_0e^{ct}\frac{-1}{c}e^{-c\tau}\Big|_0^t=$$
$$=q_0e^{ct}\frac{-1}{c}(e^{-ct}-1)=\frac{q_0}{c}(e^{ct}-1). \quad (8.9.2)$$

Нетрудно убедиться в том, что это решение удовлетворяет уравнению

$$\frac{dn}{dt}=\frac{d}{dt}\left[\frac{q_0}{c}(e^{ct}-1)\right]=q_0e^{ct}=cn+q_0$$

и условию $n=0$ при $t=0$.

Одна и та же формула (2) дает решение при положительном и отрицательном значениях c. Однако вид кривой $n=n(t)$ при этом существенно меняется. При $c>0$ (т. е. при $a>b$) показатель ct положителен, так что с ростом t величина e^{ct} быстро становится гораздо больше единицы. При большом t и $c>0$

$$n\approx\frac{q_0}{c}e^{ct}.$$

При $c<0$ величина $ct<0$; поэтому e^{ct} с ростом t становится гораздо меньше единицы и значения n приближаются к числу $-\frac{q_0}{c}$, которое положительно, так как $c<0$. Вид кривых $n=n(t)$ показан на рис. 1.

Отметим любопытный частный случай $c=0$. Если $c=0$, то формулой (2) непосредственно пользоваться нельзя. Разложим e^{ct} в ряд:

$$e^{ct}=1+ct+\frac{(ct)^2}{2}+\ldots$$

Подставляя это значение в (2), получим

$$n(t)=\frac{q_0}{c}\left[1+ct+\frac{(ct)^2}{2}+\ldots-1\right]=$$
$$=q_0\left[t+\frac{1}{2}ct^2+\ldots\right].$$

Последней формулой можно пользоваться и при $c=0$; мы получаем

$$n(t)=q_0t. \quad (8.9.3)$$

Результат (3) нетрудно вывести и из уравнения (1). Действительно, при $c=0$ уравнение (1) имеет вид $\frac{dn}{dt}=q_0$. Отсюда $n(t)=q_0t+A$, где A — постоянная интегрирования. При $t=0$ должно быть $n=0$, поэтому $A=0$, и мы получаем (3).

Как было показано выше, при $c<0$ концентрация нейтронов с течением времени достигает постоянного значения $\frac{q_0}{c}$, или, что то же самое, $\frac{q_0}{|c|}$. Чем меньше $|c|$ (чем мы ближе к критическому состоянию), тем больше это постоянное значение. Таким образом, даже при очень слабом источнике (малом q_0) масса, близкая к критической, может давать любое большое число нейтронов, большое число делений, большое выделение энергии. В принципе режим работы атомных котлов именно такой.

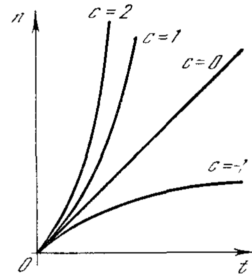


Рис. 8.9.1

Поддержание такого режима является нелегкой задачей, так как малые изменения b и c очень резко меняют величину $\frac{q_0}{c}$, когда c близко к нулю, а работа при c, близких к нулю, необходима, если надо получать большую мощность при малом q_0. Однако при помощи автоматического регулирования эту техническую задачу удается решить: когда n выходит из данных пределов, система регулирования меняет величины a или b. К тому же есть природные факторы, облегчающие регулирование. Так, например, при увеличении n повышается температура активного вещества, и при этом оказывается, что c уменьшается, так что в известных пределах система как бы саморегулируется.

§ 10. Значение критической массы

Мы знаем теперь, насколько сильно меняются свойства системы в зависимости от того, имеем ли мы дело с надкритической или с подкритической массой. Рассмотрим подробнее само условие критичности

$$r_{кр}=\frac{3k}{N\sigma_д(\nu-1)}.$$

Подставим значения входящих в последнюю формулу величин для урана-235:

$\sigma_д \approx 1,6 \cdot 10^{-24}$ см^2, $\nu \approx 2,5$, $N \approx 4 \cdot 10^{22}$ 1/см^3. Получим (в см)

$$r_{кр} \approx k \frac{3}{4 \cdot 10^{22} \cdot 1,6 \cdot 10^{-24} \cdot 1,5} \approx 30k.$$

Коэффициент k мы не умеем определять, нам известно только, что он меньше единицы. Найдем этот коэффициент, сопоставляя формулу с опытом. Опыт показывает, что критическая масса урана-235 около 50 кг. Урановый шар массой 50 кг имеет радиус около 8,5 см. Следовательно, в данном случае

$$k \approx \frac{8,5}{30} \approx 0,3.$$

Остановимся на физическом смысле формулы для критического радиуса. В выражении для $r_{кр}$ скорость нейтронов сократилась. Значит, формулу для $r_{кр}$ можно получить, не рассматривая хода процесса во времени, не рассматривая скорости размножения нейтронов и скорости ухода нейтронов из системы.

Если пренебречь безразмерным множителем $3k$ (он порядка единицы), то формула для критического радиуса принимает вид

$$r_{кр} N \sigma_д \approx \frac{1}{\nu - 1}. \qquad (8.10.1)$$

Что за величина стоит в левой части формулы (1)? Объем цилиндра с высотой, равной радиусу, и с площадью основания, равной $\sigma_д$, есть $r_{кр}\sigma_д$. Напомним, что если нейтрон движется по оси такого цилиндра, то он вызывает деление тех ядер урана-235, центры которых находятся внутри цилиндра. N — число ядер в единице объема, следовательно, $N r_{кр} \sigma_д$ — среднее число ядер в объеме цилиндра.

Теперь мы можем дать другую формулировку условию критичности. Мы выяснили, что средний путь, который проходит внутри делящегося вещества нейтрон, родившийся внутри вещества (в результате деления), порядка радиуса r. После того как нейтрон прошел путь около r, он выходит за пределы делящегося вещества и для процесса потерян. Условие критичности заключается в том, что в среднем на этом пути нейтрон должен до ухода из системы породить один нейтрон. При делении образуется $\nu - 1$ новых нейтронов. Значит, нужно, чтобы до вылета нейтрон вызвал примерно $\frac{1}{\nu - 1}$ делений, т. е. чтобы в объеме цилиндра $r\sigma_д$ было примерно $\frac{1}{\nu - 1}$ ядер. Это условие и приводит к формуле (1).

Понятно, что такие рассуждения нестроги, но они нужны для понимания сути дела, и здесь их нельзя заменить никакими расчетами, даже самыми точными вычислениями на современных электронных машинах. Такие расчеты на машинах не заменяют, а дополняют ясное понимание качественной стороны, понимание физического смысла процесса. В частности, читатель должен обратить внимание на принцип, который был высказан в начале параграфа: если какая-то величина v участвовала в выводе формулы, но в окончательном результате сократилась, значит, есть такой вывод формулы, в котором данная величина вообще не фигурирует. Нужно всегда стараться найти такой, более простой вывод формулы, потому что другой вывод формулы — это и другой, новый взгляд на исследуемое явление.

Глава 9

МЕХАНИКА

§ 1. Сила, работа, мощность

Соотношения между главнейшими механическими величинами можно точно выразить только при помощи интегралов и производных. В гл. 2 мы уже рассматривали связь между пройденным путем (или положением) тела z и его скоростью v, а также между скоростью v и ускорением a: $v=\frac{dz}{dt}$, $a=\frac{dv}{dt}$. Обратимся теперь к соотношениям, связывающим между собой такие величины, как сила, работа, энергия, мощность.

Рассмотрим прямолинейное движение тела вдоль оси x. Пусть сила F, действующая на тело, также направлена вдоль оси x. В элементарных курсах физики работа A, совершенная силой, определяется как произведение силы F на пройденный телом путь $l=b-a$, где a — начальное положение тела, а b — его конечное положение:

$$A=Fl=F(b-a).$$

Очевидно, здесь повторяется положение, с которым мы встретились в случае соотношения между скоростью и путем: простая формула — работа равна произведению силы на путь — имеет место лишь в том случае, когда сила п о с т о я н н а. Если же сила на протяжении процесса перемещения тела м е н я е т с я, то весь процесс перемещения приходится разбивать на ряд отдельных этапов так, чтобы на протяжении каждого отдельного этапа силу можно было считать постоянной (это будет выполняться, если каждому этапу отвечает малое приращение времени или пути). Тогда для малого промежутка Δx_i пути, отвечающего i-му этапу (от положения x_i до x_{i+1} тела), работа равна

$$\Delta A_i=F_i\Delta x_i=F_i(x_{i+1}-x_i)$$

Отсюда следует, что в общем случае переменной силы $F=F(x)$ работа выражается не произведением, а и н т е г р а л о м:

$$A=\int_a^b F dx.$$

Мы считаем известным движение тела, задаваемое функцией $x=x(t)$. Перемещение dx тела за малое время равно произведению (мгновенной) скорости v на время dt:

$$dx=vdt=\frac{dx}{dt}dt.$$

Поэтому выражение для работы можно переписать так:

$$A=\int_\alpha^\beta F\frac{dx}{dt}dt=\int_\alpha^\beta Fvdt, \qquad (9.1.1)$$

где моменты времени $t=\alpha$ и $t=\beta$ отвечают началу и концу движения тела.

Произведение Fv, которое входит в эту формулу, есть работа, отнесенная к единице времени; оно называется *мощностью*. Действительно, в случае постоянных скорости и силы путь равен $x=vt$, работа равна $A=Fx=Fvt$ и отношение работы к протекшему времени (т. е. работа, произведенная в единицу времени, или мощность) есть $\frac{A}{t}=Fv$. Обозначая мощность $\frac{A}{t}$ через W, можно написать

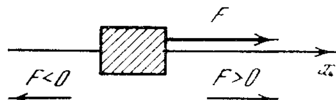
$$A=\int_\alpha^\beta Wdt. \qquad (9.1.1а)$$

Напомним, что в системе единиц СИ единица скорости есть м/с, единица ускорения — м/с^2. Единица силы имеет специальное название — *ньютон* (обозначение: Н): это есть сила, придающая массе 1 кг ускорение, равное 1 м/с^2. Ясно, что единица энергии, или работы, есть 1 Н·м=1 кг·м^2/с^2; она называется *джоуль* (обозначение: Дж); единица мощ-

ности — 1 Н·м/с=1 кг·м²/с³ — называется *ватт* (обозначение: Вт).

На тело могут действовать одновременно н е с к о л ь к о сил, например F_1 и F_2. В этом случае можно говорить о работе A_1, которую совершила первая сила, и о работе A_2 второй силы за то время, когда тело переместилось из начального положения a в конечное положение b. Считая силы F_1 и F_2 постоянными, получим:

Рис. 9.1.1



$$A_1 = (b-a)F_1, \quad A_2 = (b-a)F_2.$$

Обратим внимание на знаки величин, входящих в эти выражения. Сила считается *положительной*, если она действует в сторону возрастания x (рис. 1). Сила, действующая в противоположную сторону — влево, считается *отрицательной*. Если тело перемещается в ту же сторону, в которую действует сила, то работа силы положительна. Если же тело движется в направлении, противоположном направлению силы, так что F_1 и $(b-a)$ имеют разные знаки, то работа силы A оказывается отрицательной. Представим себе, например, что на тело действуют две силы (рис. 2, *а*): сила растянутой пружины F_1 и сила натяжения веревки F_2, которую Вы, читатель, натягиваете рукой. Сила пружины тянет тело влево, т. е. $F_1 < 0$; Вы тянете вправо, т. е. $F_2 > 0$. Если Вы тянете сильнее пружины (т. е. если абсолютная величина силы, с которой Вы тянете вправо, больше абсолютной величины силы, с которой пружина тянет тело влево: $|F_2| > |F_1|$), то тело,

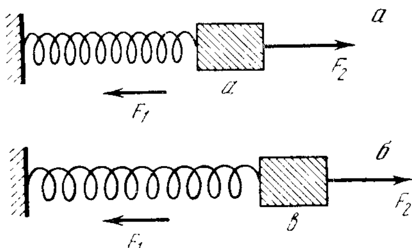


Рис. 9.1.2

первоначально находившееся в покое, будет перемещаться слева направо. На рис. 2, *а* показано начальное положение тела, а на рис. 2, *б* — его конечное положение; $(b-a) > 0$, $F_1 < 0$. Работа A_1, совершенная силой натяжения пружины, или, короче, работа пружины при этом перемещении отрицательна, а работа, которую Вы произвели, положительна: $A_2 > 0$. Полная работа $A = A_2 + A_1$ также положительна; однако $A < A_2$, так как $A_1 < 0$. Это значит, что только часть A произведенной Вами работы (A_2) воспринята телом: другая часть ($|A_1|$) пошла на растяжение пружины. Отметим, что сила трения о неподвижную поверхность направлена всегда в сторону, п р о т и в о п о л о ж н у ю скорости движения тела; поэтому работа силы трения о неподвижную поверхность всегда *отрицательна* независимо от направления движения тела.

Сила F_1 действия на тело пружины, один конец которой закреплен, отличается весьма важным свойством: эта сила зависит исключительно от положения тела. Вовсе не все силы обладают подобным свойством. Например, сила трения о неподвижную поверхность всегда тормозит движение тела: она направлена влево, если тело движется вправо, и направление вправо, если тело движется влево. Поэтому направление силы трения зависит от направления движения тела. Кроме того, сила трения может зависеть от скорости тела. Следовательно, сила трения зависит от величины и направления скорости тела, а не только от его положения (она может даже вовсе не зависеть от положения тела).

Сила F_2, с которой Вы тянете веревку, в примере, проиллюстрированном рис. 2, может меняться любым образом, по Вашему произволу. В частности, тело может двигаться сначала направо, а потом налево. При этом тело дважды пройдет одно и то же положение: первый раз на пути вперед в момент t_1, второй раз на обратном пути в момент t_2.

Возможный график движения тела, т. е. зависимость координаты x от времени t, показан для этого случая на рис. 3. В нашей воле в момент t_1 тянуть тело вправо, $F_2(t_1) > 0$, а в момент t_2 отпустить трос, так что $F_2(t_2) = 0$ или даже толкать тело влево, так что $F_2(t_2) < 0$. Но $x(t_1) = x(t_2) = x_1$; значит, произвольную силу F_2 нельзя, вообще

говоря, рассматривать как функцию координаты x.

Приведенные примеры силы трения и силы, приложенной человеком, действующим по своей воле, служат для того, чтобы показать, что зависимость силы только от положения тела $F_1=F_1(x)$, характерная для силы F_1, с которой пружина действует на тело, не есть общее свойство любых сил, а есть частное свойство, связанное с упругостью пружины.

Для того чтобы найти работу A_i данной силы F_i (где индекс i показывает, что речь идет об одной из действующих на тело сил $F_1, F_2, \ldots$), надо воспользоваться формулой $A_i=\int\limits_a^b F_i dx$ или формулой $A_i=\int\limits_\alpha^\beta F_i v dt$. При этом нужно знать: 1) как двигалось тело, т. е. зависимость $x(t)$ координаты тела от времени; 2) силу $F_i=F_i(x, t, v)$ (в общем случае сила зависит от x, t и v).

Зная зависимости $x(t)$ и $v(t)$ и подставляя их в выражение $F_i(x, t, v)$, мы получим выражение для F_i в виде функции времени и сможем записать работу как интеграл по времени. Напомним, что на тело могут действовать и другие силы, поэтому каждую отдельную силу F_i нельзя выразить через ускорение и массу тела.

П р и м е р. Пусть сила $F(x)=-kx$, а движение тела задано уравнением $x=B\sin\omega t$, т. е. $F(x,t)=-kB\sin\omega t$, и $v=\frac{dx}{dt}=B\omega\cos\omega t$; при этом здесь мы не требуем, чтобы сила F равнялась массе, умноженной на ускорение: предполагается, что на тело действуют еще и другие силы G, которые вместе с F обеспечивают движение по заданному закону $x=x(t)$. Для работы одной только силы F легко найдем

$$A=-B^2k\omega\int\limits_{t_1}^{t_2}\sin\omega t\cos\omega t dt=$$

$$=-\frac{B^2k\omega}{2}\int\limits_{t_1}^{t_2}\sin 2\omega t dt=\frac{B^2k}{4}\cos 2\omega t\Big|_{t_1}^{t_2}=$$

$$=\frac{B^2k}{4}(\cos 2\omega t_2-\cos 2\omega t_1) \qquad (9.1.2)$$

(проверьте это).

В случае, когда сила зависит только от координаты, гораздо проще и удобнее пользоваться выражением работы как интеграла по x:

$$A=\int\limits_a^b F(x)\,dx=-k\int\limits_a^b x dx=\frac{ka^2}{2}-\frac{kb^2}{2}.$$

Подставляя сюда $x=B\sin\omega t$, легко получить и выражение работы в течение заданного промежутка времени от t_1 до t_2:

$$A=\frac{kB^2\sin^2\omega t_1}{2}-\frac{kB^2\sin^2\omega t_2}{2}. \qquad (9.1.3)$$

Это выражение в точности совпадает с предыдущим: ведь

$$\cos 2\omega t=\cos^2\omega t-\sin^2\omega t=1-2\sin^2\omega t,$$

и, значит,

$$\cos 2\omega t_2-\cos 2\omega t_1=1-2\sin^2\omega t_2-$$
$$-(1-2\sin^2\omega t_1)=2(\sin^2\omega t_1-\sin^2\omega t_2).$$

Подставляя последнее тождество в (2), получим (3).

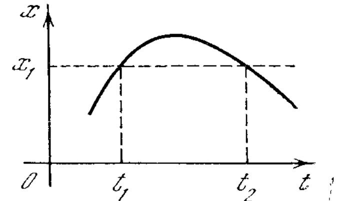


Рис. 9.1.3

При использовании выражения работы как интеграла по координате x силы $F(x, v, t)$, в общем случае зависящей от x, v, t, нужна большая осторожность. В самом деле, в принципе если задан закон движения $x=x(t)$, то уравнение $x=x(t)$ можно решить относительно t и определить $t(x)$. Однако надо иметь в виду, что t может быть н е о д н о з н а ч н о й функцией x: одно и то же положение x может отвечать, например, д в у м различным моментам времени, т. е. для одного и того же значения x мы можем иметь два разных значения t (см. рис. 3). Тогда все движение надо разбить на отдельные периоды, в течение каждого из которых скорость не меняет своего знака и t является о д н о з н а ч н о й функцией x. Однако для разных участков времени t может выражаться разными функциями x.

Пусть, например, тело движется по закону $x=B\sin\omega t$, как и в предыдущем

примере, но сила задана как функция времени: $F=f\cos\omega t$. При этом сила не является однозначной функцией положения x: ведь, если $t=0$, то $x=0$, $F=f$, но если положить $t=\frac{\pi}{\omega}$, то мы снова получим $x=0$, но на этот раз $F=-f$, так что в различные моменты времени $\left(t=0 \text{ и } t=\frac{\pi}{\omega}\right)$ тело находится в одном и том же положении $x=0$, а сила при этом различна. При интегрировании по времени этой трудности нет: каждому моменту t отвечает одно определенное значение координаты x, силы F и всех других величин.

В данном случае легко найти работу, интегрируя по времени:

$$A=\int_{t_1}^{t_2} Fvdt=\int_{t_1}^{t_2} f\cos\omega t B\omega\cos\omega t dt=$$

$$=fB\omega\int_{t_1}^{t_2}\cos^2\omega t dt.$$

Воспользуемся формулой $\cos 2\varphi=2\cos^2\varphi-1$, или, что то же самое, $\cos^2\varphi=\frac{1}{2}+\frac{\cos 2\varphi}{2}$. Тогда

$$A=fB\omega\int_{t_1}^{t_2}\left(\frac{1}{2}+\frac{\cos 2\omega t}{2}\right)dt=$$

$$=\frac{1}{2}fB\omega(t_2-t_1)+$$

$$+\frac{fB}{4}(\sin 2\omega t_2-\sin 2\omega t_1). \qquad (9.1.4)$$

Движение по закону $x=B\sin\omega t$ представляет собой колебание тела (см. гл. 10). Как видно из формулы (4), в нашем случае с течением времени, т. е. с увеличением t_2, работа неограниченно возрастает. Это объясняется тем, что сила и скорость колебаний находятся в *резонансе* (подробно явление резонанса будет рассмотрено в следующей главе).

Рассмотрим работу силы за один полупериод, выбрав начальный момент $t_1=0$, $x_1=0$ и конечный момент $t_2=\frac{\pi}{\omega}$, $\sin\omega t_2=\sin\pi=0$, $x_2=0$. Тогда в (4) $\sin 2\omega t_2=\sin 2\omega t_1=0$, и работа равна

$$A=\frac{1}{2}fB\omega\frac{\pi}{\omega}=\frac{\pi}{2}fB. \qquad (9.1.5)$$

Тело вернулось в исходное состояние, а работа, произведенная силой, не равна нулю и имеет вполне определенную величину. Как понять этот результат с точки зрения формулы $A=\int_{x_1}^{x_2} Fdx$? Если подставить $x_1=x_2=0$, то мы получим

$$A=\int_0^0 Fdx=0.$$

Но в действительности мы должны рассмотреть отдельно процесс нарастания x от 0 до $x_{max}=B$ и процесс убывания x от $x_{max}=B$ до 0. В процессе нарастания каждому значению x соответствует определенное значение силы F, которое обозначим F_1:

$$F_1=f\cos\omega t=f\sqrt{1-\sin^2\omega t}=$$

$$=f\sqrt{1-\left(\frac{x}{B}\right)^2}>0.$$

Во время убывания x тем же положительным значениям x соответствует (о т р и ц а т е л ь н о е) значение силы[1]. Эту отрицательную силу мы обозначим через F_2:

$$F_2(x)=-f\sqrt{1-\left(\frac{x}{B}\right)^2}.$$

Таким образом, интеграл с координатой x в качестве переменной интегрирования распадается на два интеграла:

$$A=\int_0^B F_1(x)\,dx+\int_B^0 F_2(x)\,dx, \qquad (9.1.6)$$

которые нельзя сложить по формуле

$$\int_a^b\varphi dx+\int_b^c\varphi dx=\int_a^c\varphi dx,$$

так как подынтегральная функция в двух интегралах правой части (6), хотя и имеет один и тот же смысл (сила), но выражается разными формулами.

[1] Равенство $\cos^2\omega t+\sin^2\omega t=1$ выполняется при л ю б ы х значениях ωt. Из него следует $\cos\omega t=\pm\sqrt{1-\sin^2\omega t}$, а вот знак перед корнем зависит от того, какова величина ωt. Легко убедиться, что при $-\frac{\pi}{2}<\omega<\frac{\pi}{2}$ надо брать знак плюс, а при $\frac{\pi}{2}<\omega t<\frac{3\pi}{2}$ — знак минус, как мы и поступали.

Это связано с тем, что F задана как функция t, а t при возрастании x от 0 до B и при убывании x от B до 0 выражается через x различными формулами. В данном случае $F_2(x) = -F_1(x)$. Подставляя выражения для $F_1(x)$ и $F_2(x)$ в (6), получим

$$A = f\int_0^B \sqrt{1-\left(\frac{x}{B}\right)^2}\,dx - $$

$$-f\int_B^0 \sqrt{1-\left(\frac{x}{B}\right)^2}\,dx.$$

Во втором интеграле можно переставить пределы интегрирования; при этом меняется знак, — и мы получим

$$A = 2f\int_0^B \sqrt{1-\left(\frac{x}{B}\right)^2}\,dx. \qquad (9.1.7)$$

Если обозначить $z = \frac{x}{B}$, то $dx = B\,dz$ и

$$A = 2Bf\int_0^1 \sqrt{1-z^2}\,dz.$$

Но интеграл

$$I = \int_0^1 \sqrt{1-z^2}\,dz = \frac{\pi}{4}$$

(этот интеграл выражает площадь четверти круга радиуса 1). Поэтому из (7) следует

$$A = 2Bf\frac{\pi}{4} = \frac{\pi}{2}Bf,$$

что совпадает с формулой (5), полученной интегрированием по времени.

Таким образом, в случае силы, зависящей от времени и могущей принимать разные значения при одинаковом значении x, работа A также не является однозначной функцией x. В разобранном случае зависимости силы от времени ($F = f\cos\omega t$, $x = B\sin\omega t$) с течением времени x снова и снова проходит одни и те же значения, а работа, произведенная силой при положительном f, все увеличивается и увеличивается.

Если сила зависит от скорости, как это имеет место для силы трения, положение дела будет похожим на разобранное нами: тело может вернуться в исходную позицию, но работа силы не будет равна нулю. В случае силы трения работа будет отрицательна (см. упражнения).

Упражнения

9.1.1. Найдите выражение в виде интеграла для работы силы трения, пропорциональной скорости движения тела и направленной в противоположную ей сторону: $F = -hv$, где $h > 0$. Покажите, что эта работа отрицательна.

9.1.2. Сила трения постоянна по величине и направлена в сторону, противоположную скорости, т. е. $F = -h$ при $v > 0$ и $F = h$ при $v < 0$. Тело движется по закону $x = B\sin\omega t$. Найдите работу силы трения за время от $t = 0$ до $t = \frac{\pi}{\omega}$.

9.1.3. Сила, действующая на тело, задана формулой $F = f_0\sin\omega_0 t$, где $f_0 = \mathrm{const}$. Ввиду того что на тело действуют также и другие силы, оно движется по закону $x = B\sin\omega_1 t$. Определите работу силы F за время от $t = 0$ до $t = T$. Рассмотрите отдельно случай $\omega_0 = \omega_1$.

9.1.4. Тело падает по закону $x = \frac{gt^2}{2}$ (ось x направлена вниз). Найдите формулу для работы силы сопротивления воздуха $F = -aS\rho\frac{v^2}{2}$, где a — коэффициент пропорциональности, зависящий от формы тела (ср. ниже § 9.14, 9.15), S — площадь поперечного сечения тела (в см²), ρ — плотность воздуха ($\approx 1{,}3\cdot 10^{-3}$ г/см³), v — скорость падения тела (в см/с). Найдите также формулу для работы силы тяжести $F = mg$, где m — масса тела.

Произведите вычисления и сравните результаты для деревянного шарика диаметром 1 см, где $a = 0{,}8$, и для стальной пули длиной 3 см, диаметром 0,7 см, где $a = 0{,}2$, полагая $t = 1$ с, 10 с, 100 с.

[Замечание. Смысл расчета состоит в том, что мы полагаем силу сопротивления воздуха малой по сравнению с силой тяжести и заметно не меняющей закон свободного падения тела. Подсчитывая работу силы сопротивления воздуха и сравнивая ее с работой силы тяжести, мы проверяем правильность исходного предположения о малой роли силы сопротивления воздуха. Точное решение задачи о падении тела с учетом сопротивления воздуха дано в § 14.]

9.1.5. Ветер, дующий со скоростью v_0, действует на парус с силой F, равной $aS\rho\frac{(v_0-v)^2}{2}$ при $v < v_0$ и $-aS\rho\frac{(v_0-v)^2}{2}$ при $v > v_0$; здесь v — скорость движения судна, S — площадь паруса, ρ — плотность воз-

духа, a — безразмерный коэффициент (для паруса, поставленного перпендикулярно к направлению ветра, $a \approx 1$). Найдите работу силы ветра при перемещении судна на b метров. Найдите мощность силы ветра. (Движение судна можно считать равномерным, т. е. скорость v постоянной.) Выясните зависимость работы и мощности от величины v. Найдите максимальную мощность при $v_0=30$ м/с, $a=1$, $S=100$ м^2 и выразите ее в ваттах.

9.1.6. Тело движется по закону $x=B\cos(\omega t+\alpha)$ под действием нескольких сил, в том числе зависящей от времени силы $F=f\cos\omega t$. Найдите работу силы за время от $t=t_1$ до $t=t_2$, в частности за один период ее действия $\left(\text{от } t=0 \text{ до } t=\frac{2\pi}{\omega}\right)$. Найдите среднюю мощность силы.

§ 2. Энергия

Рассмотрим случай силы $F=F(x)$, которая зависит только от положения (координаты) тела. Как отмечалось выше, примером силы такого рода является сила, с которой на тело действует пружина, второй конец которой закреплен [1]. В таком случае выражение $A=\int\limits_{x_1}^{x_2} F dx$ можно применять без каких-либо осложнений (ср. с § 1). В частности, в этом случае, если тело двигалось сначала в одну сторону от x_1 до $x_{max}=X$, а потом в противоположную и вернулось в начальное положение, то $x_2=x_1$ и полная работа силы действительно равна нулю:

$$A=\int\limits_{x_1}^{x_2=x_1} F(x)\,dx=0.$$

Разбиение пути на участки только подтверждает этот вывод:

$$A=\int\limits_{x_1}^{X} F dx+\int\limits_{X}^{x_2} F dx=\int\limits_{x_1}^{X} F dx-\int\limits_{x_2}^{X} F dx,$$

и $A=0$ при $x_1=x_2$.

В механике *потенциальная энергия* определяется как способность произвести работу. Пружина обладает определенным запасом потенциальной энергии, зависящим от того, насколько она растянута или сжата. При неподвижном положении одного закрепленного конца потенциальная энергия пружины зависит от положения тела, к которому прикреплен подвижный конец пружины. Таким образом, потенциальная энергия $u=u(x)$ есть функция координаты x. Если в начальном положении $x=x_1$ потенциальная энергия равна $u(x_1)$, то после перемещения тела в положение x_2, когда пружина совершила работу A, равную

$$A=\int\limits_{x_1}^{x_2} F(x)\,dx,$$

оставшаяся потенциальная энергия равна $u(x_1)-A$. Таким образом,

$$u(x_2)=u(x_1)-A=u(x_1)-\int\limits_{x_1}^{x_2} F(x)\,dx. \qquad (9.2.1)$$

Надо полностью прочувствовать знак при A в этом выражении: если пружина совершает (положительную) работу, то запас способности пружины совершать работу уменьшается! Произведенная пружиной работа черпается из запаса потенциальной энергии. Поэтому произведенная работа (то, что отдала пружина) равна разности начальной и конечной энергии пружины:

$$A=u(x_1)-u(x_2).$$

Во все формулы входит р а з н о с т ь величин потенциальной энергии в двух положениях тела. Поэтому если заменить $u(x)$ на $u(x)+C$, где C — любая постоянная величина, то это ничего не изменит в физических результатах. Действительно,

$$[u(x_1)+C]-[u(x_2)+C]=u(x_1)-u(x_2).$$

Значение $u(x)=u_0$ в какой-то данной точке (обозначим ее x_0) можно выбрать произвольно. После этого в любой другой точке x значение функции $u(x)$ определяется по формуле (1), если в ней положить $x_1=x_0$, $x_2=x$:

$$u(x)=u_0-\int\limits_{x_0}^{x} F(x)\,dx. \qquad (9.2.2)$$

Так решается задача об *определении потенциальной энергии по заданной силе.*

[1] Если второй конец пружины может перемещаться, то сила, действующая на тело, будет зависеть не только от положения тела, но и от положения второго конца пружины, а следовательно, не будет удовлетворять сформулированному условию.

Можно поставить обратную задачу: *зная потенциальную энергию $u(x)$ как функцию x, найти силу $F(x)$*. Для решения этой задачи возьмем производную от обеих частей равенства (2). Производная интеграла по верхнему пределу равна подынтегральной функции, так что

$$\frac{du(x)}{dx} = -F(x). \tag{9.2.3}$$

Знак минус здесь весьма существен. Сила положительна, направлена в сторону увеличения x в том случае, если $\frac{du}{dx}$ отрицательна, т. е. если при увеличении x потенциальная энергия u уменьшается. Сила отрицательна (направлена в сторону уменьшения x), если $\frac{du}{dx} > 0$, т. е. если при увеличении x энергия u увеличивается; в этом случае, очевидно, при уменьшении x энергия u также уменьшается. Значит, сила направлена всегда в сторону у м е н ь ш е н и я потенциальной энергии!

Рассмотрим подробнее пример п р у ж и н ы. Пусть в ненатянутом состоянии пружины тело находится в начале координат (рис. 1). При оттягивании тела вправо сила упругости пружины растет пропорционально тому расстоянию, на которое оттянуто тело, и направлена влево:

$$F = -kx, \quad \text{где } k > 0. \tag{9.2.4}$$

Положим $u_0 = 0$ при $x = 0$, т. е. примем потенциальную энергию ненатянутой пружины за нуль. Получим

$$u(x) = -\int_0^x F dx = k\int_0^x x dx = k\frac{x^2}{2}. \tag{9.2.5}$$

Легко убедиться, что этому $u(x)$ по общей формуле (3) соответствует сила (4).

Рассмотрим второй пример: с и л у т я ж е с т и. Направим ось z вверх. Сила тяжести, действующая вниз, равна $-mg$, где g — ускорение свободного падения. Она не зависит от высоты z, но ведь постоянную величину тоже можно рассматривать как функцию, значения которой не меняются. Важно то, что сила тяжести не зависит от времени и скорости. Поэтому можно применить полученные выше формулы. Примем за нуль потенциальную энергию тела на уровне Земли (т. е. при $z=0$). Тогда

$$u(z) = -\int_0^z F dz = -\int_0^z (-mg)\, dz = mgz. \tag{9.2.6}$$

Потенциальная энергия линейно растет с увеличением высоты тела над поверхностью Земли.

В последнем примере мы предполагали, что расстояние z мало по сравнению с радиусом Земли. Будем теперь рассматривать снова силу притяжения, считая, что расстояния могут быть сколь угодно большими. По закону тяготения

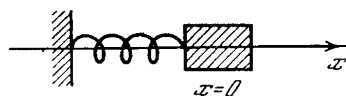


Рис. 9.2.1

Ньютона сила притяжения обратно пропорциональна квадрату расстояния от притягивающей массы. Известно, что для тела, находящегося над поверхностью Земли, сила притяжения ко всему земному шару может быть заменена силой притяжения к массе, равной массе Земли и сосредоточенной в центре Земли [2]. Удобно поэтому расстояние отсчитывать от центра Земли. Обозначим его r. Итак, сила, действующая на тело, направлена к центру Земли и по абсолютной величине равна $\frac{C}{r^2}$, где константа C положительна.

Константу C легко определить из того условия, что сила, действующая на поверхности Земли ($r = r_0 \approx 6400$ км $= 6{,}4\cdot 10^6$ м), хорошо известна: $F(r_0) = mg = \frac{C}{r_0^2}$, т. е. $C = mgr_0^2$, где g — ускорение свободного падения (на поверхности Земли $g \approx 9{,}81$ м/с^2). Окончательно

$$|F| = \frac{mgr_0^2}{r^2}. \tag{9.2.7}$$

Примем снова за нуль потенциальную энергию тела на уровне Земли. Учитывая, что при увеличении расстояния r

[2] Для тела, находящегося внутри Земли, дело обстоит уже не так: здесь при расчете силы надо учитывать только ту часть массы, которая б л и ж е к центру Земли, чем рассматриваемое тело.

от центра Земли работа силы F будет отрицательной, имеем

$$u = -\int_{r_0}^{r} F dr = mgr_0^2 \int_{r_0}^{r} \frac{dr}{r^2} =$$

$$= mgr_0^2 \left(-\frac{1}{r}\Big|_{r_0}^{r}\right) = mgr_0^2\left(-\frac{1}{r}+\frac{1}{r_0}\right) =$$

$$= mg\frac{r_0}{r}(r-r_0). \qquad (9.2.8)$$

При малой высоте $z=(r-r_0) \ll r_0$ величина $\frac{r_0}{r}$ мало отличается от единицы, и поэтому

$$u(r) \approx mg(r-r_0) = mgz, \qquad (9.2.9)$$

что совпадает с полученной раньше формулой (6). Однако с ростом r, как видно из (8), потенциальная энергия не растет неограниченно, как это было бы, если бы была верна приближенная формула (6), а стремится к определенному пределу

$$u(\infty) = mgr_0. \qquad (9.2.10)$$

Таким образом, с учетом ослабления притяжения с ростом расстояния от центра Земли энергия тела на бесконечном расстоянии будет такой же, как вычисленная по приближенной формуле энергия на расстоянии r_0 от поверхности Земли, т. е. на расстоянии $2r_0$ от центра Земли.

В этой задаче мы встречаемся с физическим вопросом, в котором речь идет о бесконечном расстоянии r. Здесь, конечно, уместно вспомнить, что в физической задаче нас могут интересовать лишь конечные величины, конечные расстояния. Так, например, если рассматриваются движение тела и энергия тела, зависящие от притяжения Земли, то непосредственно нас может интересовать процесс достижения Луны, Марса или других планет либо даже других звезд, т. е. достижение расстояний больших или очень больших по сравнению с радиусом Земли, но все же конечных! И запись $r=\infty$, разумеется, для физика может иметь лишь смысл $r \gg r_0$.

Пусть рассматривается вопрос о запуске ракеты на большую высоту, на большое расстояние от Земли. Мы хотим знать необходимую энергию и время полета. Рассмотрим два случая:

1) ракета должна пролететь расстояние $R=10r_0$, где r_0 — радиус Земли;

2) ракета должна пролететь расстояние $R=100r_0$.

Работа, необходимая для того, чтобы, оторвавшись от Земли, ракета могла удалиться на расстояние R от центра Земли, равна

$$A = mgr_0^2\left(\frac{1}{r_0}-\frac{1}{R}\right). \qquad (9.2.11)$$

Вспоминая, что $r_0 \approx 6{,}4\cdot 10^6$ м, получим в первом случае $A_1 \approx mg\cdot 5{,}76\cdot 10^6$, а во втором случае $A_2 \approx mg\cdot 6{,}34\cdot 10^6$.

Изменение расстояния в 10 раз незначительно повлияло на величину необходимой энергии. Если бы мы подставили «значение» $R=\infty$ (напоминаем, что ∞ — это не число!), то получили бы

$$A_\infty \approx mg\cdot 6{,}4\cdot 10^6.$$

Работа A_1 отличается от A_∞ на 10%, A_2 от A_∞ — на 1%, т. е. при подсчете работы A_2 (и даже работы A_1) величину R можно свободно заменить бесконечностью. Однако изменение R существенно скажется на времени полета, поэтому при рассмотрении времени заменять R бесконечностью никак нельзя.

Таким образом, одна и та же величина R в одной и той же задаче при рассмотрении разных сторон вопроса иногда может заменяться на бесконечность, иногда же такая замена не является законной. Возможность подобной замены зависит не только от самой величины R (и ее сравнения с другими входящими в формулу величинами той же размерности, в данном случае с r_0), но и от строения той формулы, в которой производится замена.

Возвращаясь к вопросу о потенциальной энергии тела, притягиваемого к Земле, найдем численное значение $u(\infty)$ на единицу массы: оно равно $gr_0 \approx 9{,}81\cdot 6{,}4\cdot 10^6 \approx 6{,}28\cdot 10^4$ Дж/г. Эта величина в 30 раз больше теплоты испарения воды и в 10 раз больше химической энергии взрывчатых веществ.

В задачах небесной механики и в физике целесообразно выбирать за нуль потенциальную энергию тела, находящегося на бесконечном расстоянии от притягивающей его массы. Тогда для потенциальной энергии тела, находящегося на расстоянии r от центра Земли, получим

$$u(r) = u(\infty) - \int_{\infty}^{r} F(r)\,dr = -\frac{C}{r}, \qquad (9.2.12)$$

где, как и ранее, C — константа в выражении силы $F=-\frac{C}{r^2}$; она определяется по формуле $C=mgr_0^2$, если известно ускорение g на поверхности Земли и радиус r_0 Земли.

Можно получить другое выражение для C. Действительно, по закону тяготения Ньютона $F=-\varkappa\frac{mM}{r^2}$, где m — масса тела, притягиваемого к Земле, M — масса Земли, r — расстояние до центра Земли, $\varkappa$ — гравитационная постоянная, $\varkappa=6{,}7\cdot 10^{-11}$ Н·м²/кг² $=6{,}7\cdot 10^{-11}$ м³/(кг·с²). Поэтому $C=\varkappa mM$. Пользуясь этой формулой, легко определить C, зная $\varkappa$ и M.

В действительности с помощью лабораторных опытов, измеряя притяжение друг к другу тяжелых шаров известной массы, находят величину $\varkappa$. Только после этого, измеряя притяжение тела к Земле, удается найти массу Земли. Вид функции, выражающей закон Ньютона, т. е. тот факт, что сила убывает обратно пропорционально квадрату расстояния, доказывается сравнением притяжения к Земле тела, находящегося на поверхности Земли, и притяжения далекого тела — Луны, а также сравнением притяжения к Солнцу планет, находящихся на различных расстояниях от Солнца: от Меркурия до Плутона.

Задача о потенциальной энергии *двух электрических зарядов* e_1 и e_2 совершенно аналогична предыдущей. Сила взаимодействия между ними равна

$$F=k\frac{e_1e_2}{r^2}. \qquad (9.2.13)$$

При этом коэффициент k в (13) равен единице, если заряды выражены в электростатической системе единиц (единица заряда $\frac{1}{3\cdot 10^9}$ Кл (кулона)), а сила — в динах (обозначение: дин; 1 дина $=10^{-5}$ Н); в системе СИ, где заряды выражаются в Кл, а сила — в Н, в формуле (13) $k=\frac{1}{9}\cdot 10^{-13}$. В формуле нет знака минус, который стоит в выражении для силы тяготения. Действительно, если заряды e_1 и e_2 одноименные (оба положительны или оба отрицательны), то произведение e_1e_2 положительно. Но в этом случае заряды отталкиваются, т. е. сила F положительна.

Определяя снова $u(r)$ так, чтобы было $u(\infty)=0$, получим

$$u(r)=k\frac{e_1e_2}{r}. \qquad (9.2.14)$$

Потенциальная энергия двух одноименных зарядов на конечном расстоянии положительна: они отталкиваются и, расходясь с расстояния r до ∞, могут совершить работу, равную

$$u(r)-u(\infty)=u(r).$$

Потенциальная энергия двух разноименных зарядов отрицательна. Действительно, $e_1e_2<0$, если $e_1>0$, $e_2<0$. Это ясно и из физических соображений: так как разноименные заряды притягиваются, то нужно затратить энергию для того, чтобы растащить их на бесконечное расстояние.

Отметим, что благодаря закону сохранения энергии потенциальную энергию можно определить не только как способность производить работу, но и как работу, которую нужно было затратить для приведения системы в данное состояние. Растянутая пружина способна произвести определенную работу, возвращаясь в нерастянутое состояние. Очевидно, что именно такую же работу надо было затратить для того, чтобы растянуть пружину. Аналогичные утверждения можно высказать в случае тела, поднятого на определенную высоту над Землей, или для системы двух зарядов.

§ 3. Равновесие и устойчивость

Рассмотрим тело, которое может двигаться без трения вдоль какой-то прямой, принимаемой нами за ось x. Пусть на тело действует сила, направленная вдоль этой оси и зависящая от координаты x (здесь можно снова представить себе, например, *пружину*; ниже мы разберем и другие примеры).

Положение *равновесия* характеризуется равенством нулю действующей на тело силы, т. е. тем, что тело в этом положении может оставаться в покое. Точку, в которой имеет место равновесие, обозначим через x_0, так что $F(x_0)=0$. Разлагая функцию $F(x)$ в ряд Тейлора и пренебрегая всеми степенями $(x-x_0)$, кроме первой, видим, что возможны два варианта зависимости $F=F(x)$ в окрестности точки x_0 равновесия, где $F(x_0)=0$: $F(x)\approx k_1(x-x_0)$ и

$F(x) \approx -k_2(x - x_0)$ (здесь мы считаем, что k_1 и k_2 — положительные величины). Первый случай показан на рис. 1, *а*, а второй — на рис. 1, *б*.

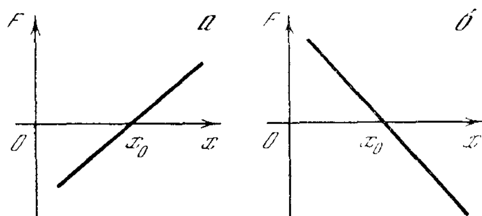


Рис. 9.3.1

Этим случаям соответствует совершенно различный характер равновесия. В самом деле, в случае рис. 1, *а*, если тело находится несколько правее точки $x = x_0$, то на него действует положительная сила, т. е. сила, которая тянет его дальше вправо. Таким образом, в случае рис. 1, *а* равновесие в точке $x = x_0$ *неустойчивое*: достаточно малого отклонения тела (безразлично вправо или влево), чтобы на тело начала действовать сила, у в е л и ч и в а ю щ а я это отклонение. Напротив, в случае рис. 1, *б* при отклонении тела вправо сила отри-

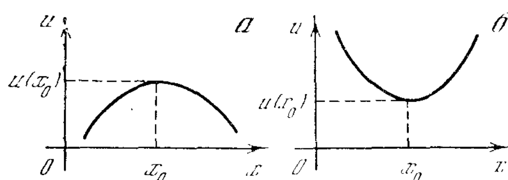


Рис. 9.3.2

цательна, т. е. тянет влево. Отклонение тела от положения равновесия вызывает силу, стремящуюся в о з в р а т и т ь тело в положение равновесия. В этом случае имеем дело с *устойчивым* равновесием. Легко убедиться, что для тела, закрепленного на пружине, осуществляется второй случай.

В соответствии с выписанными выше выражениями силы находим выражения потенциальной энергии, пользуясь (2.2). В случае неустойчивого равновесия

$$u(x) \approx u(x_0) - \frac{1}{2} k_1 (x - x_0)^2.$$

В случае устойчивого равновесия

$$u(x) \approx u(x_0) + \frac{1}{2} k_2 (x - x_0)^2.$$

Графики функции $u = u(x)$ для этих двух случаев изображены на рис. 2.

Таким образом, в случае неустойчивого равновесия потенциальная энергия имеет м а к с и м у м, в случае устойчивого равновесия — м и н и м у м. В обоих случаях в самой точке максимума или минимума $F = -\left.\frac{du}{dx}\right|_{x=x_0} = 0$, т. е. сила равна нулю.

Этот результат является вполне естественным. Если тело находится в состоянии максимума потенциальной энергии, то при перемещении в обоих направлениях выделяется энергия, которая, превращаясь в кинетическую энергию, может пойти на преодоление инерции. Если же тело находится в состоянии минимума энергии, то для перемещения его в любое другое положение нужно затратить энергию извне; эта энергия пойдет на увеличение потенциальной энергии; затратив малую энергию, можно будет лишь незначительно сместить тело. Такие свойства тела, находящегося в положении минимума потенциальной энергии, полностью соответствуют понятию устойчивого равновесия.

При действии силы тяжести вблизи поверхности Земли потенциальная энергия равна mgz, где z — высота над поверхностью Земли. Линии, изображающие зависимость $u(x)$, здесь совпадают с графиком зависимости высоты z тела от горизонтальной координаты x. Надо представить себе тело, которое движется вдоль кривой линии как бусина, нанизанная на жесткую проволоку. Кривая $u(x)$ соответствует форме проволоки, если плоскость чертежа расположить вертикально. Тогда очевидно, что максимум $u(x)$ (см. рис. 2, *а*) отвечает точке проволоки, с которой бусины при малейшем толчке скатываются вниз, уходя вправо и влево от максимума, минимум же $u(x)$ (см. рис. 2, *б*) — нижней точке, в которой бусины находятся в устойчивом положении и куда они сами скатываются с соседних участков.

Таким образом, по графику $u(x)$ можно наглядно представить себе направление сил и характер равновесия.

Рассмотрим несколько примеров.

П р и м е р 1. Пусть заряженное тело движется по прямой (примем ее за ось x), на которой закреплены неподвижно два одинаковых заряда e_1, сим-

метричных относительно начала координат и удаленных на расстояние $2a$ один от другого (рис. 3).

Ясно, что в этом случае начало координат является положением равновесия тела. Действительно, здесь силы, действующие на тело со стороны неподвижных зарядов, равны по величине и противоположны по направлению, так что они уравновешиваются, т. е. их равнодействующая равна нулю.

Потенциальная энергия $u(x)$ тела с зарядом e равна

$$u(x)=\frac{e_1e}{r'}+\frac{e_1e}{r''},$$

где r' и r'' — соответственно расстояния от тела до левого и правого зарядов, т. е. $r'=x+a$, $r''=a-x$. Поэтому

$$u(x)=e_1e\left(\frac{1}{a+x}+\frac{1}{a-x}\right). \qquad (9.3.1)$$

Соответствующие кривые изображены на рис. 4. Верхняя кривая соответствует случаю $e_1e>0$, т. е. случаю одноименных зарядов тела и неподвижных зарядов, нижняя — случаю $e_1e<0$, т.е. случаю, когда тело имеет заряд, противоположный по знаку неподвижным зарядам.

В случае $e_1e<0$ равновесие в начале координат неустойчиво. Действительно, тело притягивается и к левому, и к правому заряду; в начале координат силы притяжения к двум зарядам уравновешивают друг друга. Однако если тело немного сдвинется, например, вправо, то притяжение к правому заряду окажется сильнее и тело будет продолжать двигаться в ту же сторону. Аналогичные соображения убеждают нас и в том, что при $e_1e>0$ равновесие устойчиво: здесь при малом сдвиге тела вправо (большее) отталкивание правого заряда будет стремиться вернуть его в исходное положение.

Найдем $\left.\frac{d^2u}{dx^2}\right|_{x=0}$. Пользуясь (1), получаем

$$\frac{d^2u}{dx^2}=2e_1e\left[\frac{1}{(a+x)^3}+\frac{1}{(a-x)^3}\right]. \qquad (9.3.2)$$

Полагая в формуле (2) $x=0$, находим

$$\left.\frac{d^2u}{dx^2}\right|_{x=0}=\frac{4e_1e}{a^3}.$$

Следовательно, если $e_1e>0$, то $\frac{d^2u}{dx^2}>0$ при $x=0$. В этом случае $u(x)$ имеет минимум при $x=0$, равновесие устойчиво. Если же $e_1e<0$, то

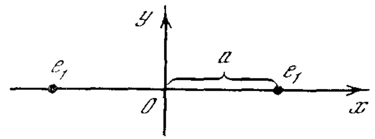


Рис. 9.3.3

$\left.\frac{d^2u}{dx^2}\right|_{x=0}<0$, т. е. $u(x)$ имеет максимум при $x=0$, равновесие неустойчиво.

Пример 2. Рассмотрим случай, когда заряды расположены на том же

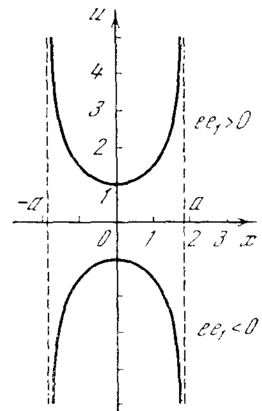


Рис. 9.3.4

расстоянии от начала координат, но на прямой, перпендикулярной к линии (оси абсцисс), по которой движется заряженное тело (рис. 5). В этом случае потенциальная энергия равна

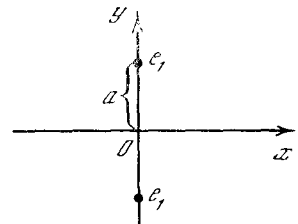


Рис. 9.3.5

$$u(x)=2\frac{e_1e}{\sqrt{a^2+x^2}}.$$

График функции $u(x)$ при $a=1$, $|e_1e|=1$ изображен на рис. 6. При $e_1e>0$ равновесие в начале координат неустойчиво. При заряде e тела, отличающемся по знаку от неподвижных зарядов e_1, т. е. при $e_1e<0$, равновесие устойчиво.

Это нетрудно установить, рассматривая силу, действующую на подвижный заряд (рис. 7). Пусть $e_1e > 0$. Сместим тело вправо от положения равновесия. Тогда равнодействующая сил отталкивания направлена тоже вправо, в сторону дальнейшего увеличения отклонения. Равновесие неустойчивое.

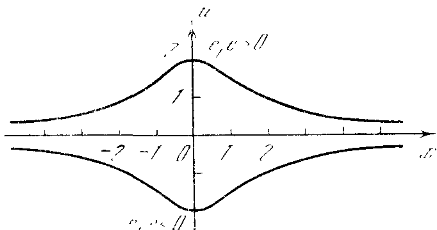


Рис. 9.3.6

В случае $e_1e < 0$ равнодействующая направлена в сторону уменьшения отклонения; равновесие устойчивое.

К этим результатам также нетрудно прийти, рассматривая вторую производную $\left.\frac{d^2u}{dx^2}\right|_{x=0}$ (сделайте это сами).

Обратим внимание, что при $e_1e > 0$, когда в примере 1 (см. рис. 3) была устойчивость, в примере 2 (см. рис. 5) равновесие неустойчиво. При $e_1e < 0$ (разноименные заряды) положение обратное: равновесие неустойчиво при расположении зарядов в соответствии с рис. 3 и устойчиво при расположении их, как на рис. 5.

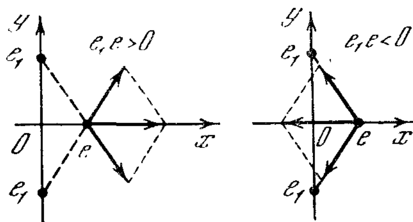


Рис. 9.3.7

Повернув рис. 5 на 90°, мы заметим, что он, в сущности, относится к тому же начальному распределению зарядов в положении равновесия, что и рис. 3. Можно сказать, что рис. 3 и 5 относятся к одинаковому начальному распределению зарядов, но рассматриваются разные направления движения. При этом всегда (при любых знаках зарядов) в каком-либо из направлений движения равновесие оказывается неустойчивым.

В электростатике доказывается, что этот результат является совершенно общим: в пространстве между внешними закрепленными зарядами нигде нет такой точки равновесия, чтобы равновесие было устойчивым относительно перемещений в любом направлении.

Приведенное ниже общее доказательство этого факта может показаться читателю довольно сложным; его можно пропустить без ущерба для понимания дальнейшего материала.

Заметим, что зависимость потенциальной энергии заряда e, находящегося в точке (x, y, z), от его расстояния r от неподвижного заряда e_1, помещенного в точку (x_1, y_1, z_1), дается формулой

$$u = \frac{e_1e}{r} = \frac{e_1e}{[(x-x_1)^2+(y-y_1)^2+(z-z_1)^2]^{1/2}}.$$

Рассмотрим движение по оси x и найдем $\frac{d^2u}{dx^2}$ при постоянных y и z; эта величина называется (второй) частной производной от u по x и обозначается через $\frac{\partial^2u}{\partial x^2}$. Затем аналогично найдем $\frac{\partial^2u}{\partial y^2}$ и $\frac{\partial^2u}{\partial z^2}$, относящиеся соответственно к движению по осям y и z. Оказывается[1], что в нашем случае при любых x, y, z, x_1, y_1, z_1 *сумма вторых производных по трем переменным* x, y, z (им отвечают три взаимно перпендикулярные оси координат) *равна нулю*[2]:

$$\frac{\partial^2u}{\partial x^2}+\frac{\partial^2u}{\partial y^2}+\frac{\partial^2u}{\partial z^2}=0. \qquad (9.3.3)$$

Очевидно, это свойство сохранится и для суммы любого числа членов вида $\frac{e_ke}{r_k}$, где e_k — неподвижный заряд в точке (x_k, y_k, z_k), а r_k — расстояние заряда e от этой точки. Следовательно, при любом распределении неподвижных зарядов в пространстве имеет место формула (3).

[1] Читателю следует самому в этом убедиться. Заметим, что точка (x_1, y_1, z_1), в которой находится неподвижный заряд, отличается тем, что в ней функция u и ее производные бесконечны; эту точку мы исключаем из рассмотрения. Относительно путей, позволяющих рассматривать такие точки, см. гл. 16.

[2] Функции $u(x, y, z)$, для которых выполняется это условие, называют *гармоническими*; они играют большую роль во многих разделах физики (ср. гл. 15 и 17).

В частности, соотношение (3) справедливо и для той точки, где имеет место положение равновесия заряда e. Но для равновесия необходимо, чтобы в этой точке сумма проекций сил на каждую из осей была равна нулю. Для этого нужно, чтобы

$$\frac{\partial u}{\partial x}=0, \quad \frac{\partial u}{\partial y}=0, \quad \frac{\partial u}{\partial z}=0$$

— ведь если проекции силы на три перпендикулярные оси равны нулю, то равна нулю и сама сила (векторная величина!), т. е. равна нулю проекция силы на любое направление [3].

Для того чтобы равновесие было устойчивым относительно движения по всем трем взаимно перпендикулярным направлениям, нужно, чтобы выполнялись неравенства:

$$\frac{\partial^2 u}{\partial x^2}>0, \quad \frac{\partial^2 u}{\partial y^2}>0, \quad \frac{\partial^2 u}{\partial z^2}>0. \qquad (9.3.4)$$

Однако (4) противоречит (3), так как сумма трех положительных величин не может равняться нулю.

Упражнения

9.3.1. Заряд e движется по прямой, на которой закреплены неподвижно два положительных заряда e_1 и $e_2=4e_1$ на расстоянии $2a$ один от другого. Найдите ту точку на прямой, в которой возможно равновесие заряда e, и установите характер равновесия. Рассмотрите отдельно два случая: $e>0$ и $e<0$.

9.3.2. Решите упр. 1, изменив знак заряда e_2.

§ 4. Второй закон Ньютона

Сформулированный Ньютоном второй закон движения тела заключается в том, что *произведение массы на ускорение равно действующей силе* [1]. Ускорение a есть производная скорости v по времени; в свою очередь скорость есть производная от координаты тела по времени. Таким образом,

$$ma=m\frac{dv}{dt}=F, \qquad (9.4.1)$$

или

$$m\frac{d^2x}{dt^2}=F. \qquad (9.4.2)$$

Начнем со случая, когда сила задана как функция времени: $F=F(t)$. Это означает, что производная $\frac{d^2x}{dt^2}$ задана как функция времени. В таком случае из закона Ньютона (1) легко найти скорость в любой заданный момент; кроме действующей силы при этом надо задать также скорость в какой-то момент t_0. Тогда

$$v(t)=v(t_0)+\frac{1}{m}\int_{t_0}^{t} F(t)\,dt. \qquad (9.4.3)$$

Зная скорость в зависимости от времени $v(t)$ и начальное положение тела $x(t_0)$, найдем положение тела в любой момент времени:

$$x(t)=x(t_0)+\int_{t_0}^{t} v(t)\,dt, \qquad (9.4.4)$$

где $v(t)$ дано предыдущей формулой.

Соотношение между скоростью и путем подробно рассматривалось в гл. 2.

В целом формулы (3) и (4) решают задачу нахождения $x(t)$ из уравнения (2). Соотношение (2) — это *дифференциальное уравнение второго порядка*: в него входит вторая производная неизвестной функции $x(t)$. В ответ входит не только заданная функция $x(t)$, но и две постоянные величины, определяемые из начальных условий, которыми служат положение тела и скорость тела в начальный момент времени t_0.

Если задан или экспериментально найден закон движения тела, т. е. задана функция $x(t)$, то нахождение силы, действующей на тело, не представляет никакого труда: надо найти вторую производную функции $x(t)$ и умножить ее на m (формула (2)).

Упражнения

9.4.1. Найдите закон движения тела под действием постоянной силы F, если в момент времени $t=0$ тело покоилось в начале координат ($x=0$).

9.4.2. То же, при условии: а) $x=0$, $v=v_0$ при $t=0$; б) $x=x_0$, $v=v_0$ при $t=0$.

9.4.3. Тело массой 20 кг начинает двигаться под действием силы 1 Н из начала координат без начальной скорости. Какой путь пройдет оно за 10 с?

---

[3] Если имеется отличная от нуля сила F (вектор!), действующая в каком-либо направлении, то вдоль каждой оси будет действовать сила, равная проекции силы F на эту ось.

[1] Первый закон Ньютона — закон инерции — утверждает, что *тело, на которое не действуют никакие силы, движется прямолинейно и равномерно*. Это значит, что при равной нулю силе ускорение тела также равно нулю. Следовательно, первый закон Ньютона содержится во втором законе, является его частным случаем.

9.4.4. Шарик свободно, т. е. с начальной скоростью, равной нулю, падает с высоты 100 м. Через сколько времени он достигает Земли (сопротивлением воздуха пренебречь)?

9.4.5. В условиях предыдущей задачи шарик начинает падение со скоростью $v_0=10$ м/с. Разберите два случая: а) начальная скорость шарика v_0 направлена вниз; б) начальная скорость шарика v_0 направлена вверх. Определите, через сколько времени шарик достигнет Земли. Какова его скорость в момент приземления? Проверьте, что скорость приземления одинакова в случаях а) и б).

9.4.6. На тело действует сила, пропорциональная времени, протекшему с начала движения (коэффициент пропорциональности равен k). Найдите закон движения тела, если известно, что тело начинает движение из точки $x=0$ с начальной скоростью v_0.

9.4.7. На тело действует сила, периодически меняющаяся со временем: $F=f\cos\omega t$, где f и ω — постоянные числа.

а) Найдите закон движения тела при условии, что $x=0$, $v=0$ при $t=0$. Установите, что такое движение является колебательным. Определите период колебания, наибольшее значение $x(t)$, наибольшее значение скорости.

б) Тот же вопрос для силы $F=f\sin\omega t$, причем снова $x=0$, $v=0$ при $t=0$.

9.4.8. Тело движется под действием постоянной силы F. В момент времени $t=t_0$ тело находилось в точке $x=x_0$. Найдите, какую скорость должно иметь тело при $t=t_0$, чтобы в момент времени $t=t_1$ попасть в точку $x=x_1$.

§ 5. Импульс силы

Задача нахождения закона движения тела при заданной зависимости силы от времени в принципе решена в предыдущем параграфе. Проанализируем полученное решение и введем некоторые связанные с ним новые понятия.

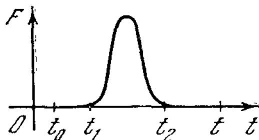


Рис. 9.5.1

Произведение $P=mv$ (массы на скорость) называется *количеством движения*. Величина

$$I(t_0,\ t)=\int_{t_0}^{t} F(t)\,dt \qquad (9.5.1)$$

называется *импульсом силы* за время от t_0 до t.

Формула (4.3) может быть записана так:

$$P(t)-P(t_0)=\int_{t_0}^{t} F dt\ \ (=I(t_0,\ t)) \qquad (9.5.2)$$

— изменение количества движения равно импульсу силы.

Часто встречается случай силы, действующей только в течение краткого промежутка времени; примером является удар молотком, отскакивающим после удара от тела. И до, и после удара сила равна нулю. Если отсутствуют другие силы, кроме краткой силы удара, то очевидно, что до удара тело движется с постоянной скоростью и после удара оно движется с другой, но также постоянной скоростью.

Пусть $F(t)$ отлично от нуля только в промежутке от t_1 до t_2 (рис. 1). Рассмотрим интеграл

$$I=\int_{t_1}^{t_2} F(t)\,dt. \qquad (9.5.3)$$

Эту величину можно назвать полным импульсом силы — полным в том смысле, что интеграл берется по всему промежутку времени, когда сила фактически действует.

В выражение (1) входит интеграл от t_0 до t. Если $t_0<t_1$, а $t>t_2$, то

$$I(t_0,\ t)=\int_{t_0}^{t} F dt=I.$$

Действительно, запишем

$$\int_{t_0}^{t} F dt=\int_{t_0}^{t_1} F dt+\int_{t_1}^{t_2} F dt+\int_{t_2}^{t} F dt.$$

В правой части первый и третий интегралы равны нулю, так как в соответствующих промежутках $F=0$, а второй (средний) интеграл есть I. Таким образом, из формул (2) и (3) получаем $P(t)=P(t_0)+I$, если $t_0<t_1$, $t>t_2$.

Из формулы (4.3) следует, что скорость после удара зависит только от импульса силы, т. е. от интеграла силы; конкретный вид функции $F(t)$ на ней не сказывается. Например, несколько различных кривых $F(t)$, нарисованных на рис. 2, *а*, дают одинаковый импульс силы, т. е. одинаково

меняют скорость тела. Для каждой из линий $F(t)$ нетрудно нарисовать соответствующий график скорости $v(t)$. На рис. 2, *б* изображены такие графики в общем предположении, что начальная скорость равна нулю. Общим для всех кривых рис. 2, *б* является конечное значение скорости: все кривые переходят справа в горизонтальную прямую на высоте $v=\frac{I}{m}$.

Каждую из кривых $F(t)$ рис. 2 можно сжимать по оси времени и пропорционально растягивать по оси силы. Площадь под кривой F, т. е. $\int F dt$ — полный импульс силы, при этом не изменяется. Именно так получена, например, линия *2* на рис. 2, *а* из линии *1*.

Чем короче время действия силы, тем меньше время, в течение которого скорость тела меняется от начального значения $v_0=0$ до конечного значения $v=\frac{I}{m}$ (см. рис. 2, *б*). Таким образом, в пределе при очень большой силе, действующей в течение очень малого времени, график скорости приобретает вид ступеньки (ср. рис. 3, *а* с кривой *2* на рис. 2, *б*). При этом становится несущественным, какую из кривых рис. 2, *а* мы сжимали: ступенька характеризуется только одной величиной $v=\frac{I}{m}$, а она для всех кривых одинакова.

Если до действия силы тело покоилось в точке x_0, то после кратковременного действия большой силы тело начинает двигаться с постоянной скоростью, равной $\frac{I}{m}$. Если сила действовала в момент t_1 (промежуток времени от t_1 до t_2 мы считаем малым и не различаем t_2 и t_1), то положение тела в зависимости от времени дается формулами:

$$x=x_0 \quad \text{при } t<t_1,$$
$$x=x_0+\frac{I}{m}(t-t_1) \quad \text{при } t>t_1. \tag{9.5.4}$$

Соответствующий график показан на рис. 3, *б*. Отметим, что $x(t)$ удовлетворяет уравнению

$$m\frac{d^2x}{dt^2}=F(t).$$

Напомним, что на графике (x, t) первая производная $\frac{dx}{dt}$ связана с наклоном касательной к кривой. Вторая производная характеризует скорость изменения первой производной, т. е. вторая производная связана с

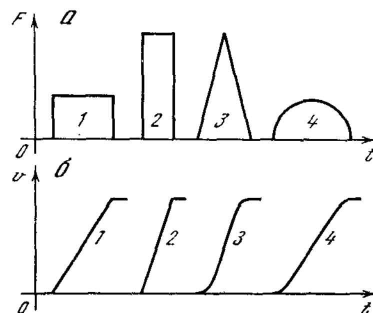


Рис. 9.5.2

кривизной линии $x(t)$ (см. § 7.9).

На рис. 3, *б* линия $x(t)$ имеет излом в точке $t=t_1$, $x=x_0$. Излом можно представить себе как точку, в которой кривизна бесконечна, так что наличие излома соответствует рассмо-

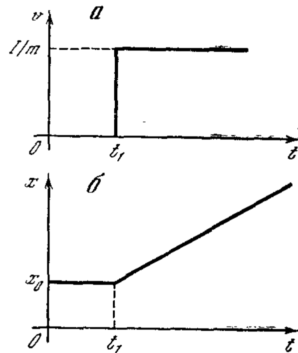


Рис. 9.5.3

трению очень большой (в пределе — бесконечной) силы. Однако до и после излома производная $\frac{dx}{dt}$ конечна: значит, очень большая сила действовала очень малое время, так что импульс силы конечен [1]. Импульс легко определить по графику (см. рис. 3, *б*), вычисляя скорость после действия силы и используя формулу (2).

Найденный нами закон движения тела, которое покоилось до момента $t=\tau$, а в этот момент получило импульс

[1] Подробнее о связанных с этим примером математических конструкциях (δ-функции Дирака) говорится в гл. 16.

силы I, поможет нам усовершенствовать формулы (4.3) и (4.4) предыдущего параграфа. Для этого нам понадобится специальный случай формулы (4), когда при $t=\tau$ тело находилось в начале координат ($x_0=0$). Введем специальное обозначение:

$$x_1(t, \tau)=\begin{cases} 0 & \text{при } t<\tau, \\ \frac{I}{m}(t-\tau) & \text{при } t>\tau. \end{cases} \qquad (9.5.5)$$

Если подставить $v(t)$ из (4.3) в формулу (4.4) и использовать более аккуратные обозначения (так, чтобы верхнему пределу и переменной интегрирования отвечали разные буквы), то получится выражение, на первый взгляд довольно громоздкое:

$$x(t)=x(t_0)+(t-t_0)v(t_0)+$$
$$+\frac{1}{m}\int_{t_0}^{t} dt_1 \int_{t_0}^{t_1} F(t_2)\,dt_2. \qquad (9.5.6)$$

Стоящий здесь интеграл можно преобразовать по формальным правилам обращения с повторными интегралами. Однако мы нигде не упоминали об этих правилах, и поэтому интересующее нас преобразованное выражение в виде однократного интеграла мы получим, пользуясь законом (5) движения тела под действием одиночного импульса силы.

Действие силы $F(t)$ за интервал времени $\Delta\tau$ от некоторого момента τ до $\tau+\Delta\tau$ можно приближенно заменить (мгновенным) действием импульса силы: $\Delta I=F(\tau)\,\Delta\tau$. Движение тела под действием такого импульса нам уже известно (см. формулу (5), в которой лишь нужно заменить I на $\Delta I(\tau)$).

Дальше остается только сложить вклады всех интервалов Δt_i в координату $x(t)$; в результате мы получим

$$x(t)\approx\sum x_1(t, \tau)=\sum\frac{1}{m}(t-\tau)F(\tau)\Delta\tau\approx$$
$$\approx\frac{1}{m}\int_{t_0}^{t} F(\tau)(t-\tau)\,d\tau. \qquad (9.5.7)$$

Здесь, как обычно, мы заменили сумму большого числа членов, отвечающих достаточно малым интервалам $\Delta\tau$, на интеграл (причем левую и правую части (7), отбросив промежуточные части, можно соединить знаком строгого равенства). Формула (7) не учитывает начальной координаты $x(t_0)$, а также движения $(t-t_0)\,v(t_0)$ под действием начальной скорости, ибо мы в (7) считали, что такое движение отсутствует ($v(t_0)=0$, как и $x(t_0)=0$). Более общее выражение получим, если прибавим эти члены к левой части (7):

$$x(t)=x(t_0)+(t-t_0)v(t_0)+$$
$$+\frac{1}{m}\int_{t_0}^{t} F(\tau)(t-\tau)\,d\tau. \qquad (9.5.8)$$

Формула (8) выгодно отличается от (6) тем, что в (8) интегрировать нужно только один раз. Мы не сказали, почему можно просто складывать слагаемые, отвечающие отдельным импульсам, начальной скорости и начальной координате. Подробнее об этом говорится в § 17.4. Здесь же нам достаточно того, что можно непосредственно проверить, чему равны $x(t_0)$, $\left.\frac{dx}{dt}\right|_{t_0}$ и $\frac{d^2x}{dt^2}$, исходя из формулы (8). Для этого нужно продифференцировать даваемое формулой (8) $x(t)$ аналогично тому, как это делалось при проверке формулы (8.3.5).

Напомним читателю, что по третьему закону Ньютона при взаимодействии двух тел сила F_1, с которой второе тело действует на первое, равна по величине и противоположна по направлению силе F_2, с которой первое тело действует на второе [2]:

$$F_2(t)=-F_1(t).$$

Формула (2) в применении к первому телу и силе F_1 дает

$$P_1(t)-P_1(t_0)=\int_{t_0}^{t} F_1 dt. \qquad (9.5.9)$$

Эта же формула в применении ко второму телу и силе F_2 дает

$$P_2(t)-P_2(t_0)=\int_{t_0}^{t} F_2 dt. \qquad (9.5.10)$$

Так как по третьему закону Ньютона $F_2=-F_1$, то

$$\int_{t_0}^{t} F_2 dt=-\int_{t_0}^{t} F_1 dt.$$

---

[2] Здесь индекс у силы F указывает, на какое тело действует эта сила; индекс у буквы P также указывает номер тела, к которому относится это количество движения.

Поэтому (10) принимает вид

$$P_2(t) - P_2(t_0) = -\int_{t_0}^{t} F_1 dt. \qquad (9.5.11)$$

Сопоставляя (9) и (11), находим

$$P_1(t) - P_1(t_0) = P_2(t_0) - P_2(t),$$

отсюда

$$P_1(t) + P_2(t) = P_1(t_0) + P_2(t_0).$$

Последняя формула показывает, что *действие тел друг на друга не меняет суммы количества движения этих тел.*

§ 6. Кинетическая энергия

Рассмотрим тело, движущееся под действием известной силы $F(t)$, и найдем связь между работой, произведенной этой силой, и скоростью движения тела.

Умножая обе части основного уравнения (второй закон Ньютона) $m\frac{dv}{dt} = F(t)$ на скорость v, получим

$$mv\frac{dv}{dt} = F(t)v. \qquad (9.6.1)$$

Но согласно правилу вычисления производных от сложных функций (см. § 4.3) $v\frac{dv}{dt} = \frac{d}{dt}\left(\frac{v^2}{2}\right)$ при любой зависимости $v = v(t)$. Поэтому (1) можно переписать так:

$$m\frac{d}{dt}\left(\frac{v^2}{2}\right) = F(t)v,$$

или, поскольку m — постоянная величина,

$$\frac{d}{dt}\left(\frac{mv^2}{2}\right) = F(t)v.$$

Вводя обозначение

$$\frac{mv^2}{2} = K, \qquad (9.6.2)$$

получим окончательно:

$$\frac{dK}{dt} = F(t)v. \qquad (9.6.3)$$

Вспомнив выражение (1.1) для работы, запишем

$$A = \int_{t_0}^{t_1} F(t)v dt = \int_{t_0}^{t_1} \frac{dK}{dt} dt,$$

откуда

$$A = K(t_1) - K(t_0). \qquad (9.6.4)$$

Величина K есть *кинетическая энергия* тела; формула (4) выражает **закон сохранения энергии**: *изменение кинетической энергии тела равно работе, произведенной силой.* Формула (3) выражает закон: *скорость изменения кинетической энергии равна мощности, которую развивает сила.*

Когда сила задана определенной функцией времени, то импульс силы и, следовательно, вызванное данной силой изменение количества движения тела не зависят ни от массы тела, ни от его начальной скорости, так как импульс силы и изменение количества движения есть $\int_{t_0}^{t_1} F dt$. Напротив, работа силы и изменение кинетической энергии тела под действием силы, как видно из формул (2)—(4), существенно зависят не только от самой силы, но и от массы тела и его начальной скорости. В самом деле, действуя заданной силой в течение определенного времени на тяжелое тело, покоившееся в начале движения, мы придадим ему лишь малую скорость: перемещение тела будет мало и работа силы также будет мала. Легкое тело заберет на себя большую работу, приобретет большую энергию. Если до начала действия силы тело двигалось в сторону, противоположную направлению действия силы, то сила может уменьшить его энергию.

§ 7. Инерциальные и неинерциальные системы отсчета

Представим себе тело, участвующее одновременно в двух движениях. Например, человек ходит по каюте парохода, или в каюте падает брошенный мяч, а пароход движется. Предположим, что одно из этих движений (в нашем примере — движение парохода) равномерное. Возникает вопрос: можно ли, наблюдая падение мяча в каюте либо движение в ней какого-нибудь тела под действием приложенной силы, установить, движется пароход или нет? Иначе говоря, влияет ли равномерное движение парохода на характер движения предметов на самом пароходе? Оказывается, что нет, никак не влияет. Опыты показали, что отсутствие влияния равномерного движения на физические явления относится не только

к механике, но и к процессу распространения света, к электрическим и магнитным явлениям. Из этого факта Эйнштейн сумел сделать выводы огромной важности, развив специальную, а затем и общую теорию относительности (в нашей книге теория относительности не излагается). Ньютон, формулируя законы движения, предполагал, что существует *абсолютное время*, общее для всего мира, и что также существует *абсолютное пространство*, в котором развертываются все мировые процессы. В предыдущем изложении мы молчаливо принимали эту гипотезу Ньютона. Можно представить себе, что где-то помещено начало координат O, задающееся, скажем, исходящими из точки O тремя взаимно перпендикулярными металлическими прутами с делениями (тремя осями координат); в этой же точке расположены и часы. Тогда все просто и однозначно: координаты материального тела определяются как проекции его положения на прутья (на оси координат) [1]; тело покоится, если его координаты постоянны. В общем случае, зная зависимость $x(t)$, $y(t)$, $z(t)$ координат от времени, можно найти скорость тела и его ускорение (точнее, три составляющих скорости $\frac{dx}{dt}$, $\frac{dy}{dt}$, $\frac{dz}{dt}$ и три составляющих ускорения $\frac{d^2x}{dt^2}$, $\frac{d^2y}{dt^2}$, $\frac{d^2z}{dt^2}$). Опыт (в данном случае астрономические наблюдения) показывает, что законы Ньютона справедливы в системе координат, начало которой совмещено с центром тяжести солнечной системы, одна ось направлена на Полярную звезду, вторая — на какую-нибудь подходящую звезду на экваторе и третья — также на звезду, находящуюся на оси, перпендикулярной первым двум осям. Масса Солнца равна $2\cdot10^{30}$ кг, масса Юпитера — $1{,}9\cdot10^{27}$ кг, а его расстояние от Солнца — $777{,}8\times\times10^9$ м; масса других планет гораздо меньше массы Юпитера. Поэтому центр тяжести солнечной системы находится на расстоянии около $0{,}78\cdot10^9$ м от центра Солнца со стороны, обращенной к Юпитеру. Радиус Солнца равен $7\cdot10^8$ м, так что центр тяжести солнечной системы, а следовательно, и «начало» координат, которым пользуются в астрономических расчетах, лежат вне поверхности Солнца.

Можно в качестве системы координат воспользоваться земными ориентирами. Например, поместим начало координат на Красной площади в Москве, одну ось направим вертикально вверх, вторую — в горизонтальной плоскости на север, третью — по широте, по направлению запад—восток. В такой системе законы Ньютона выполняются с меньшей точностью; о причинах этого будет сказано ниже.

Даже без специальных опытов, из самих законов Ньютона следует, что система координат, в которой эти законы справедливы, не единственная. В самом деле, второй закон Ньютона имеет вид

$$m\frac{d^2x}{dt^2}=F_x. \tag{9.7.1}$$

Если добавить к x линейную функцию от времени, то ускорение не изменится: ведь если $x_1(t)=x(t)+a+bt$, то

$$\frac{dx_1}{dt}=\frac{dx}{dt}+b, \quad \frac{d^2x_1}{dt^2}=\frac{d^2x}{dt^2}=F_x.$$

Но при каком изменении системы координат положение тела, которое раньше характеризовалось функцией $x(t)$, будет задано функцией $x_1(t)=x(t)+a+\\+bt$? Если в момент $t=0$ перенести начало координат в точку O_1, причем $x(O_1,\ t=0)=-a$, то в новой системе координат в момент $t=0$ будем иметь $x_1=x+a$. Если к тому же новое начало координат движется влево со скоростью $-b$, то в момент t положение нового начала O_1 имеет координату $x(O_1)=-a-bt$. Отсчитанное от этого начала положение тела

$$x_1=x(t)+a+bt, \tag{9.7.2}$$

т. е. как раз такое, какое нам нужно, чтобы закон Ньютона остался справедливым (см. выше). Новая система движется относительно старой системы с постоянной скоростью $-b$, т. е. так же, как движется по инерции

[1] Для простоты мы везде рассматриваем тело малого размера (иногда в этом случае говорят о «материальной точке»). Теория движения твердых тел конечного размера, которые могут еще вращаться и деформироваться, слишком сложна для данной книги. Если имеются в виду тела конечных размеров, то в дальнейшем, говоря о положении и движении тела, мы всегда имеем в виду положение и перемещение его центра тяжести.

тело, на которое не действуют никакие силы. Поэтому говорят, что законы Ньютона справедливы в *инерциальных системах координат* — в системах координат, движущихся одна относительно другой по закону инерции, без участия сил.

Еще до Ньютона принцип равноправия инерциальных систем отсчета сформулировал Галилей. Он писал, что в каюте корабля, скользящего с постоянной скоростью по спокойной воде, все явления происходят в точности так же, как и в каюте покоящегося корабля. Таким образом, среди всего огромного разнообразия систем координат, которые можно связать, например, с любой звездой, планетой или кометой, выделяется один более узкий класс инерциальных систем, в которых выполняются законы Ньютона.

В системе координат, вращающейся относительно инерциальной системы, появляются дополнительные члены в уравнениях движения (центробежная и кориолисова сила, но в данной книге нет возможности подробно говорить о них). Вращение Земли вносит поправки в уравнения, записанные в земных координатах. К счастью, ориентируя оси по далеким звездам, мы с большой точностью осуществляем практически невращающуюся инерциальную систему (т. е. одну из инерциальных систем). В течение какого-то времени существовали ученые, делавшие из этого факта вывод о том, что само свойство инерции зависит от реального наличия далеких звезд; однако сейчас эта точка зрения сдана в архив. Обсуждать можно не (безусловное) свойство инерции, а лишь свойства (точнее, расположение) звезд. Ближайшие звезды удалены от Солнца на расстояния порядка 10^{16}—10^{17} м и движутся относительно Солнца со скоростями порядка 10—50 км/с. Хотя эти скорости и не малы, но угловые перемещения звезд ничтожны; поэтому-то связанную с далекими звездами (ориентированную по ним) систему координат и можно считать инерциальной системой, в которой Солнце покоится.

Можно ли отказаться от инерциальных систем? Интересна, например, ситуация в системе координат, где ориентация осей не изменяется, оси с течением времени не поворачиваются, но начало координат движется ускоренно вдоль одной из осей относительно начала координат какой-то инерциальной системы отсчета.

Итак, пусть исходная система координат — инерциальная и нам задан закон движения начала координат новой системы O_1 (x):

$$x(t) = -f(t).$$

Тогда $x_1(t) = x(t) + f(t)$ и, значит,

$$\frac{d^2x_1}{dt^2} = \frac{d^2x}{dt^2} + \frac{d^2f}{dt^2} = -\frac{1}{m}F_x + \frac{d^2f}{dt^2}.$$

Измеряя $x_1(t)$ и применяя закон Ньютона к этой величине, наблюдатель придет к выводу, что на тело, кроме известной ему силы, действует еще одна сила, равная $m\frac{d^2f}{dt^2}$. Сравнивая разные тела, наблюдатель убедится, что эту дополнительную силу нужно считать пропорциональной массе тела; лишь в этом случае дополнительное (появившееся в новой системе координат) ускорение $\frac{d^2f}{dt^2}$ не будет зависеть от массы тела.

Но свойство пропорциональности силы массе тела характерно для силы тяготения! Таким образом, явления, происходящие в неинерциальной системе координат, движущейся с некоторым ускорением, подобны тем, которые происходят под действием сил тяготения. Эти очень простые соображения послужили для Эйнштейна исходным пунктом при создании современной теории тяготения — так называемой *общей теории относительности*.

Приведем один замечательный пример ускоренно движущейся системы. Рассмотрим наблюдателя, находящегося в кабине лифта. Пока лифт покоится, наблюдатель ощущает силу земного тяготения. В частности, наблюдатель, спокойно стоящий в кабине, опирается на пол кабины с силой, равной его весу Mg, где M — масса наблюдателя, g — ускорение свободного падения.

Теперь представим себе, что в момент $t=0$ оборвались тросы, удерживающие кабину, и она начала падать по закону свободного падения:

$$\frac{d^2Z}{dt^2} = -g, \quad Z = z_0 - \frac{gt^2}{2};$$

здесь Z — высота кабины над уровнем Земли. Рассмотрим наблюдателя внутри кабины, отсчитывающего свою высоту z_1 от пола лифта. С высотой над уровнем Земли z координата z_1 связана уравнением

$$z_1 = z - Z = z - z_0 + \frac{gt^2}{2}.$$

Напишем уравнение для координат z_1 тела (материальной точки), на которое действуют какие-то силы F (натяжение пружины и т. п.) и сила тяготения, равная $-gm$. В земной системе координат уравнение движения точки будет таково:

$$m\frac{d^2z}{dt^2} = F - gm.$$

Используя это уравнение, можно найти уравнение движения точки в системе координат z_1, связанной с лифтом:

$$m\frac{d^2z_1}{dt^2} = m\left(\frac{d^2z}{dt^2} + g\right) = F.$$

В этой системе координат исчезла сила тяжести! В свободно падающем лифте имеет место невесомость. В частности, для тела, на которое никакие внешние силы не действуют, имеем $\frac{d^2z_1}{dt^2} = 0$; этому уравнению удовлетворяет решение $z_1 = \text{const}$ — состояние покоя относительно лифта. (Напомним, что с точки зрения земного наблюдателя кабина и тело в ней падают с одинаковой скоростью и одинаковым ускорением g.)

Состояние невесомости закончится лишь тогда, когда кабина, достигнув дна шахты, по необходимости остановится (может быть, разбившись при этом на куски). До этого момента во время падения невозможно опытами внутри кабины установить существование поля тяготения Земли.

Работа над настоящей книгой (в одном из предыдущих ее вариантов) была начата еще до первых полетов человека в космос. В настоящее время большинство из Вас уже смотрели телевизионные передачи с бортов космических кораблей. Вы убедились воочию, что на борту корабля имеет место явление невесомости. Но ведь космический корабль удалился от Земли всего на 200—300 км, когда в силу закона Ньютона сила тяготения уменьшилась всего на несколько процентов:

$$\frac{\Delta g}{g} = -2\frac{\Delta R}{R} = -2\frac{300}{6400} \approx -0{,}09 = -9\%.$$

Значит, та полная невесомость, которая царит на борту корабля, получена в основном не за счет ослабления силы тяжести вследствие удаления от Земли, а за счет того, что космический корабль при выключении двигателей сам находится в режиме свободного падения, подобен падающему лифту.

Теперь можно снова и по-другому взглянуть на инерциальную систему отсчета, связанную с Солнцем. Мы не утверждаем, что пренебрежимо малы и гравитационный потенциал, и сила тяготения, влияющая на Солнце и на все тела в солнечной системе вследствие притяжения других звезд нашей галактики и других галактик. При равномерной (в среднем) плотности вещества и потенциал, и сила выражаются интегралами, имеющими бесконечное значение (этот факт носит название «гравитационный парадокс»). Но абсолютное значение гравитационного потенциала нигде в теории не фигурирует; производная потенциала — сила притяжения со стороны внешних тел — также полностью компенсируется свободным «падением» солнечной системы независимо от того, конечна ли эта сила или бесконечна. Так обосновывается возможность изучения солнечной системы независимо от всей остальной Вселенной, устраняется гравитационный парадокс.

§ 8*. Преобразования Галилея. Энергия в движущейся системе отсчета

Вернемся теперь к инерциальным системам отсчета и рассмотрим подробнее соотношения между описанием одного и того же явления в различных инерциальных системах. Пусть тело как-то движется в поезде, который в свою очередь идет с постоянной скоростью v_0 в определенном (указываемом рельсами) направлении, которое мы примем за направление оси x. Будем еще считать, что поезд едет в сторону возрастания абсцисс x. В таком случае координата x_1 тела по отношению к неподвижному относительно железнодорожных путей наблюдателю — скажем, стоящему на железнодорож-

ной станции, от которой отъехал поезд, — будет связана с координатой x тела относительно выбранной внутри поезда системы отсчета (начало которой может совпасть, скажем, с задней стенкой вагона) соотношением

$$x_1 = x + v_0 t + a,$$

где величина a зависит от выбора начала отсчета координат x_1: так, если начало координат x_1 совпадает с пунктом, откуда выехал поезд, то a есть x_1-координата задней стенки вагона в момент $t=0$ (ср. с формулой (7.2)). Если, кроме того, считать начало отсчета «поездного» времени t (за такой «начальный момент» можно принять, скажем, начало движения поезда) отличным от начала отсчета «абсолютного» (не связанного с поездом) времени t_1 (например, можно положить, что момент $t_1=0$ совпадает с «нулевыми показаниями» часов в пункте, от которого мы отъезжаем [1]), то формулу для x_1 придется еще дополнить следующей:

$$t_1 = t + b,$$

где b — «абсолютное» время t_1 в момент $t=0$ (в момент отхода поезда). Преобразования

$$x_1 = x + v_0 t + a, \quad t_1 = t + b \tag{9.8.1}$$

определяют связь между координатами (x_1, t_1) и (x, t) одного и того же тела в двух инерциальных системах отсчета, одна из которых (поезд) равномерно движется относительно другой (станция) со скоростью v_0; они называются *преобразованиями Галилея*.

Предположим теперь, что тело движется в поезде (в сторону возрастания абсцисс x) с некоторой скоростью v; тогда

$$x = x_0 + vt. \tag{9.8.2}$$

При этом, очевидно, имеем

$$x_1 = x + v_0 t + a = (x_0 + vt) + v_0 t + a = (x_0 + a) + (v_0 + v)t. \tag{9.8.3}$$

(Можно также, заменив t на $t_1 - b$, переписать последнюю формулу так: $x_1 = [(x_0+a)-(v+v_0)\,b] + (v+v_0)\,t_1$; мы, однако, для простоты не будем различать t_1 и t.) Из (3) следует, что по отношению к наблюдателю, стоящему на станции, тело движется со скоростью $v_1 = v + v_0$. При этом она, разумеется, будет разной для наблюдателя, стоящего на станции, и для наблюдателя, едущего в поезде. Но вот ускорение тела для этих двух наблюдателей будет уже одинаковым:

$$a_1 = \frac{dv_1}{dt} = \frac{d}{dt}(v + v_0) = \frac{dv}{dt} + \frac{dv_0}{dt} = \frac{dv}{dt} = a$$

— ведь постоянное слагаемое $v_0 = \text{const}$ в выражении для скорости не изменит ускорения. Поэтому и сила, действующая на тело, не меняется:

$$F = ma_1 = ma.$$

Разность скоростей тела до и после действия силы тоже одинакова для наблюдателя, находящегося на станции, и для наблюдателя, едущего в поезде. Действительно, пусть скорость тела по отношению к наблюдателю, стоящему в поезде, до действия силы есть v', а после действия силы v''; для наблюдателя, находящегося вне поезда, соответствующие скорости будут v_1' и v_1''. Тогда $v_1' = v' + v_0$ и $v_1'' = v'' + v_0$, поэтому

$$v_1'' - v_1' = v'' + v_0 - v' - v_0 = v'' - v'.$$

Сложнее обстоит дело с **кинетической энергией**: не только сама кинетическая энергия, но даже и разности кинетических энергий **различны** для различных наблюдателей. Для наблюдателя, стоящего на станции,

$$K_1'' - K_1' = \frac{m(v_1'')^2}{2} - \frac{m(v_1')^2}{2} = \frac{m(v'' + v_0)^2}{2} - \frac{m(v' + v_0)^2}{2} = \frac{m(v'')^2}{2} - \frac{m(v')^2}{2} + mv_0v'' - mv_0v' = K'' - K' + mv_0(v'' - v').$$

В этой формуле K_1'', K_1' — конечная и начальная кинетическая энергия, вычисленная находящимся вне поезда наблюдателем, а K'' и K' — соответствующая кинетическая энергия, вычисленная пассажиром поезда.

Работа силы и мощность для различных наблюдателей также будут различны, так как хотя сила одна и та же, но и пройденные пути,

[1] Разумеется, при далекой поездке в новом положении поезда время 00 ч 00 мин может измениться в результате переезда в другую часовую зону, что делает особенно уместным различия «поездного времени» t и «времени пункта отправления» t_1.

и скорости будут различны для наблюдателя, стоящего на станции, и для наблюдателя в поезде. Однако, хотя изменение кинетической энергии и работа силы будут неодинаковы для разных наблюдателей, равенство этих двух величин будет иметь место для к а ж д о г о наблюдателя! (Примеры, подтверждающие этот закон, имеются в упражнениях.)

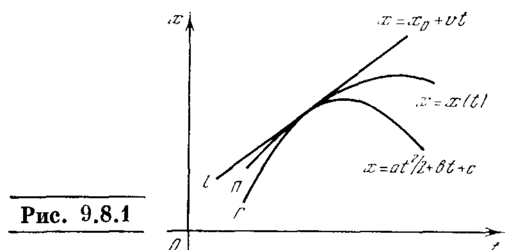


Рис. 9.8.1

Отметим замечательную формулу, которая справедлива, если тело движется под действием одной только данной силы $F(t)$:

$$A = \int_{t_0}^{t_1} F(t)\,v(t)\,dt = \frac{mv_1^2}{2} - \frac{mv_0^2}{2} =$$

$$= \frac{m}{2}(v_1 + v_0)(v_1 - v_0) =$$

$$= \frac{v_1 + v_0}{2}(mv_1 - mv_0) = \frac{v_1 + v_0}{2}\int_{t_0}^{t_1} F(t)\,dt.$$

Таким образом, в этом случае скорость $v(t)$ можно вынести из-под интеграла, заменяя ее средним арифметическим начальной и конечной скоростей движения.

Этот вывод справедлив только в том случае, если $v(t)$ есть скорость, полученная телом в результате действия лишь о д н о й силы $F(t)$. Если же на тело действует н е с к о л ь к о сил: F_1, F_2, F_3, то работа, произведенная всеми силами, равна произведению средней скорости на сумму импульсов всех сил:

$$A = \frac{v_1 + v_0}{2}\int_{t_0}^{t_1}(F_1 + F_2 + F_3)\,dt =$$

$$= \frac{v_1 + v_0}{2}\int_{t_0}^{t_1} F_1 dt + \frac{v_1 + v_0}{2}\int_{t_0}^{t_1} F_2 dt +$$

$$+ \frac{v_1 + v_0}{2}\int_{t_0}^{t_1} F_3 dt. \qquad (9.8.4)$$

Однако работа каждой из сил в отдельности (например, силы F_2) не равна соответствующему слагаемому $\frac{v_0 + v_1}{2}\int_{t_0}^{t_1} F_2 dt$ в формуле (4), так как сила F_2, действуя отдельно, сообщила бы телу скорость, отличную от скорости $v(t)$ (см. упр. 6).

Выше мы видели, что выбранные на прямой (ось x) две инерциальные системы отсчета (x, t) и (x', t') связаны одна с другой формулами

$$x' = x + vt + a, \quad t' = t + b, \qquad (9.8.5)$$

где v — скорость системы отсчета (x, t) относительно системы отсчета (x', t') (так что скорость системы отсчета (x', t') относительно начала координат (x) равна $-v$). Преобразования Галилея (5) (или (1)) играют в механике роль, сходную с той ролью, которую играют в геометрии *движения* (ср. с § 1.9). Переход от одной системы отсчета событий к другой инерциальной системе не сказывается ни на каких физических эффектах, поэтому физически значимыми являются лишь те факты и величины, запись которых в координатах (x, t) не меняет своей формы при преобразованиях (5). Так, например, л и н е й н о е уравнение

$$x = vt + a \qquad (9.8.6)$$

выражает р а в н о м е р н о е движение точки вдоль оси Ox (вдоль той самой прямой, движения по которой мы единственно и рассматриваем), однако при этом конкретное значение скорости движения v зависит от выбора системы отсчета и физического смысла не имеет. (Именно поэтому в законах Ньютона участвуют лишь у с к о р е н и я материальных частиц, но не скорости.) Напротив, если мы имеем два равномерных движения, выражаемых формулой (6), и формулой $x = v_1 t + a_1$, то разность $v_1 - v$ имеет физический смысл: это есть о т н о с и т е л ь н а я с к о р о с т ь одного движения относительно другого, не зависящая от выбора (инерциальной) системы отсчета. Физический смысл имеет интервал $\tau = t_2 - t_1$ времени между двумя событиями (x_1, t_1) и (x_2, t_2); если же эти события одновременны, т. е. $t_1 = t_2$, то можно говорить и о пространственном расстоянии $d = x_2 - x_1$ между этими событиями (при этом «расстояние» τ между событиями измеряется в единицах времени, например в секундах, а «второе расстояние» d, имеющее смысл лишь при $\tau = 0$, — совсем в других единицах, например в метрах).

Произвольная линия $x = x(t)$ плоскости x, t (рис. 1) задает закон движения (материальной) точки вдоль прямой; при этом наклон $k = x'\left(= \frac{dx}{dt}\right)$ касательной к кривой в точке (x, t) (его, пожалуй, было бы удобнее обозначать буковой v) указывает с к о р о с т ь движения. Замена самой кривой Г с уравнением $x = x(t)$ ее касательной l, приближающей

Г в окрестности рассматриваемой точки, означает переход от произвольного движения к «приближающему его» равномерному движению со скоростью, равной мгновенной скорости в рассматриваемый момент времени. Вторая производная $x''\left(=\frac{d^2x}{dt^2}\right)$ играет роль кривизны кривой; физически она выражает ускорение движения. Постоянство ускорения

$$\frac{d^2x}{dt^2}=a\ (=\text{const})$$

приводит к закону движения

$$x=\frac{at^2}{2}+bt+c, \qquad (9.8.7)$$

где b и c произвольны: таким равноускоренным движениям геометрически отвечают линии (7) — параболы плоскости x, t. Замена произвольной линии $x=x\ (t)$ приближающей ее параболой П с уравнением (7), характеризующейся теми же производными $x'\ (=v)$ и $x''\ (=a)$, означает замену произвольного движения равноускоренным с теми же (мгновенными) скоростью и ускорением; геометрически эта процедура сводится к замене кривой $x=x\ (t)$ плоскости $(x,\ t)$ соприкасающейся параболой (ср. с § 7.9). Физически переход от произвольной кривой к ее касательной означает «отключение сил», ибо прямой отвечает равномерное движение, которое происходит без участия сил (движение по инерции); замена же кривой параболой (7) равносильна предположению о постоянстве силы [2].

Упражнения

9.8.1. Найдите формулу для кинетической энергии тела, движущегося под действием постоянной силы F (в начальный момент времени скорость была равна нулю), в зависимости от времени, а также в зависимости от пройденного пути.

9.8.2. Тело движется под действием силы $F=f\cos\omega t$, причем $v=0$ при $t=0$. Найдите выражение для кинетической энергии тела. Определите максимум кинетической энергии.

9.8.3. Тело движется по закону $x\ (t)=A\cos\ (\omega t+\alpha)$; A, ω, α — постоянные числа. Определите среднюю кинетическую энергию при условии, что t неограниченно растет от значения $t=0$.

9.8.4. Шарик падает с высоты H из состояния покоя. Покажите, что кинетическая энергия шарика $K=mg\ (H-h)$, где h — высота шарика над Землей в данный момент времени.

9.8.5. Поезд массой 500 т вышел со станции и, пройдя за 3 мин 1,5 км, развил скорость 45 км/ч. Определите работу и среднюю мощность паровоза, предполагая, что: а) трения о рельсы нет; б) имеет место трение о рельсы и коэффициент трения $k=0{,}004$. [Сила трения равна силе притяжения поезда к Земле (т. е. его весу), умноженной на k.]

9.8.6. На тело действуют две силы: $F_1=at$ и $F_2=a\ (\Theta-t)$. Импульсы этих сил за время от 0 до Θ одинаковы. В момент времени $t=0$ тело имело скорость v_0. Найдите работу каждой силы за время от 0 до Θ и сравните ее с произведением импульса на среднюю скорость.

9.8.7. Человек, стоя неподвижно на Земле, действует на данную массу m силой F в течение времени t. В результате этого масса, находившаяся раньше в покое, приобрела скорость $v_1=\frac{Ft}{m}$ и кинетическую энергию $\frac{mv_1^2}{2}$, равную работе, произведенной человеком.

Рассмотрите такой же опыт, проделанный в поезде, движущемся со скоростью v_0. Масса m имела до опыта скорость v_0, а после опыта — скорость v_0+v_1. Найдите изменение кинетической энергии массы m. Какую работу произвел человек? Считая, что человек твердо опирается о стенки вагона и скорость v_0 не меняется, найдите работу силы, произведенной поездом (паровозом) во время опыта.

9.8.8. Человек массы M, стоя на коньках на льду (трением о лед пренебрегаем), действует силой F на массу m в течение времени t. Какую кинетическую энергию приобретает масса m? Какую кинетическую энергию приобретает человек? Чему равна полная работа, произведенная силой над массой m и над человеком? Почему она больше, чем в упр. 7?

9.8.9. Тот же опыт, что и в предыдущей задаче, производит человек, который в начальный момент катился со скоростью v_0 вместе с массой m. Скорость массы m после действия силы равна $v_0+\frac{Ft}{m}$, скорость человека $v_0-\frac{Ft}{M}$. Найдите изменение кинетической энергии массы m и человека в результате действия силы. Найдите работу силы, равную изменению суммарной кинетической энергии, и сравните ее с результатом предыдущего упражнения.

9.8.10. Докажите, что при преобразованиях Галилея (5) сохраняет свое значение: а) (временнóй) интервал τ между двумя собы-

[2] См.: *Яглом И. М.* Принцип относительности Галилея и неевклидова геометрия. М.: Наука, 1969, с. 97.

тиями; б) (пространственное) расстояние d между двумя одновременными событиями; в) относительная скорость v_1-v одного равномерного движения относительно другого.

9.8.11. Проверьте, что переход к новой (инерционной) системе отсчета с помощью преобразований Галилея (5) не меняет ускорения a (равноускоренного) движения (7).

§ 9*. Траектория снаряда. Парабола безопасности

Рассмотрим задачу о полете снаряда, выброшенного из орудия с начальной скоростью v_0. Примем точку вылета снаряда из ствола за начало координат, ось y направим вертикально вверх, а ось x будем считать горизонтальной (движение снаряда происходит в одной плоскости (x, y)). Для простоты силу сопротивления воздуха не будем принимать во внимание (учет этой силы значительно усложняет задачу).

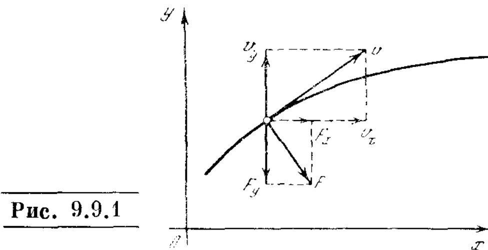


Рис. 9.9.1

По второму закону Ньютона

$$m\frac{dv}{dt}=F. \tag{9.9.1}$$

Этот закон мы применяли раньше только для прямолинейного движения. Однако в задаче о траектории направление v меняется с течением времени (скорость всегда направлена по касательной к траектории полета снаряда). Поэтому нам приходится считать, что величины v и F в формуле (1) являются *векторами*:

$\mathbf{v}=(v_x, v_y)$ и $\mathbf{F}=(F_x, F_y)$.

Всякое движение в плоскости x, y можно рассматривать как результат сложения двух движений: одного, происходящего вдоль оси x под действием силы F_x со скоростью v_x, и другого вдоль оси y под действием силы F_y со скоростью v_y. Применив к каждому из этих движений в отдельности второй закон Ньютона, получим:

$$m\frac{dv_x}{dt}=F_x,\quad m\frac{dv_y}{dt}=F_y, \tag{9.9.1a}$$

где $(v_x, v_y)=\mathbf{v}$ и $(F_x, F_y)=\mathbf{F}$ — векторы скорости и силы (рис. 1). В каждом из уравнений (1а) сила и скорость направлены вдоль одной прямой (вдоль оси x в первом уравнении и вдоль оси y во втором).

Обозначим через φ угол, образуемый стволом орудия с горизонтальным направлением; φ называют *углом бросания*. Так как мы считаем, что в процессе полета на снаряд действует только сила тяжести, направленная к Земле, то $F_x=0$, $F_y=-mg$. Поэтому уравнения (1а) в этом случае имеют вид:

$$m\frac{dv_x}{dt}=0,\quad m\frac{dv_y}{dt}=-mg. \tag{9.9.2}$$

Замечательно (и важно), что наши уравнения «разделились»: в первом из них фигурирует единственная неизвестная функция v_x, а во втором — функция v_y.

Найдем начальные условия для функций $v_x(t)$ и $v_y(t)$. В момент $t=0$ вылета снаряда из орудия

$$v_x(0)=v_0\cos\varphi,\quad v_y(0)=v_0\sin\varphi,$$

где $v_0\,(=|\mathbf{v}_0|)$ — начальная скорость снаряда. Первое из уравнений (2) дает $\frac{dv_x}{dt}=0$, откуда следует, что скорость v_x постоянна, а поэтому

$$v_x(t)=v_x(0)=v_0\cos\varphi. \tag{9.9.3}$$

Второе уравнение (2) утверждает, что $\frac{dv_y}{dt}=-g$; интегрируя обе его части в пределах от 0 до t, находим

$$v_y(t)-v_y(0)=-gt,$$

или

$$v_y(t)=-gt+v_0\sin\varphi. \tag{9.9.4}$$

Для определения перемещений x и y вдоль координатных осей воспользуемся тем, что $\mathbf{v}=\frac{d\mathbf{r}}{dt}$, где $\mathbf{r}=(x, y)$ — радиус-вектор снаряда. Это векторное уравнение расщепляется на два скалярных:

$$\frac{dx}{dt}=v_x,\quad \frac{dy}{dt}=v_y, \tag{9.9.5}$$

или в силу (3) и (4)

$$\frac{dx}{dt}=v_0\cos\varphi,\quad \frac{dy}{dt}=-gt+v_0\sin\varphi. \tag{9.9.6}$$

Уравнения (6) снова разделились: первое из них связывает неизвестную функцию $x(t)$, а второе — функцию $y(t)$.

В начальный момент времени снаряд был в начале координат, поэтому

$$x=0, \quad y=0 \quad \text{при} \quad t=0. \qquad (9.9.7)$$

Интегрируя уравнения (6) от 0 до t и пользуясь начальными условиями (7), находим:

$$x=v_0 t\cos\varphi, \qquad y=v_0 t\sin\varphi-\frac{gt^2}{2}. \qquad (9.9.8)$$

Формулы (8) дают возможность определить положение снаряда в **л ю б о й** момент времени t.

Беря различные значения t, можно найти по формулам (8) положение снаряда в различные моменты времени и построить график полета снаряда (траекторию снаряда). Таким образом, уравнения (8) дают кривую на плоскости x, y. Эти уравнения определяют параметрическое задание траектории полета снаряда, где роль параметра играет время t (см. § 1.8).

Из уравнений (8) нетрудно исключить t и получить уравнение трактории в виде зависимости y от x. Действительно, первое уравнение (8) дает $t=\frac{x}{v_0\cos\varphi}$; после этого из второго уравнения (8) находим

$$y=x\operatorname{tg}\varphi-x^2\frac{g}{2v_0^2\cos^2\varphi}. \qquad (9.9.9)$$

Из (9) видно, что y есть многочлен второй степени от x; график такого многочлена — п а р а б о л а. Поэтому траектория снаряда (если не учитывать сопротивление воздуха) есть парабола. На рис. 2 изображена траектория (9) для случая $v_0=80$ м/с, $\varphi=45°$.

Из (9) видно, что при одном и том же v_0 форма траектории зависит от угла бросания φ. Найдем *наивысшую точку траектории* снаряда при данных φ и v_0. Для определения координаты x_1 точки, в которой снаряд достигает наибольшей высоты подъема y_{max}, составим уравнение $\frac{dy(x_1)}{dx}=0$. Получаем

$$\operatorname{tg}\varphi-x_1\frac{g}{v_0^2\cos^2\varphi}=0,$$

откуда

$$x_1=\frac{v_0^2\sin\varphi\cos\varphi}{g}=v_0^2\frac{\sin 2\varphi}{2g}.$$

При этом значении $x=x_1$ высота y имеет м а к с и м у м (физически ясно, что это именно максимум; разумеется, это легко проверить и по знаку $\frac{d^2y}{dx^2}$). Подставляя найденное x в (9), получаем

$$y_{max}=y(x_1)=\frac{v_0^2\sin^2\varphi}{2g}.$$

Для определения дальности x_2 полета нужно определить то значение $x=x_2$, при котором $y(x_2)=0$ (см. рис. 2):

$$x_2\operatorname{tg}\varphi-x_2^2\frac{g}{2v_0^2\cos^2\varphi}=0.$$

Отбрасывая не интересующее нас решение $x_2=x_0=0$, находим

$$x_2=\frac{v_0^2\sin 2\varphi}{g}. \qquad (9.9.10)$$

Дальность стрельбы зависит от начальной скорости и от угла бросания.

При каком же угле бросания (при фиксированной начальной скорости v_0 снаряда) дальность стрельбы будет наибольшей? Из формулы (10) видно, что это будет тогда, когда $\sin 2\varphi=1$, т. е. при $\varphi=45°$; наибольшая дальность полета снаряда равна $\frac{v_0^2}{g}$. Кроме того, так как $\sin 2\varphi=\sin 2(90°-\varphi)$, то в силу (10) при углах наклона φ_0 и $90-\varphi_0$ (например, при углах наклона $\varphi=30°$ и $\varphi=60°$) снаряд попадает в одну и ту же

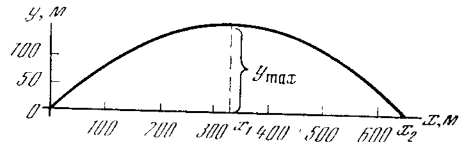


Рис. 9.9.2

точку; артиллеристы называют траекторию снаряда, отвечающую углу наклона $\varphi\leqslant 45°$, *настильной*, а отвечающую углу наклона $\varphi>45°$ — *навесной*.

Определим время t_1, в течение которого снаряд поднимается вверх. Для этого достаточно решить уравнение $\frac{dy(t_1)}{dt}=0$, потому что в тот момент $t=t_1$, когда y достигает своего наибольшего значения, снаряд перестает подниматься и начинает падать. Условие $\frac{dy}{dt}=0$ дает $v_0\sin\varphi-gt=0$, откуда

$$t_1=\frac{v_0\sin\varphi}{g}. \qquad (9.9.11)$$

Полное время полета t_2 определим исходя из того, что полет прекращается в тот момент, когда $x = x_2$. Пользуясь (8) и (10), находим

$$v_0 t_2 \cos\varphi = \frac{v_0^2 \sin 2\varphi}{g},$$

откуда

$$t_2 = \frac{2v_0 \sin\varphi}{g}. \qquad (9.9.12)$$

Сравнивая (12) и (11), видим, что при любом угле наклона орудия полное

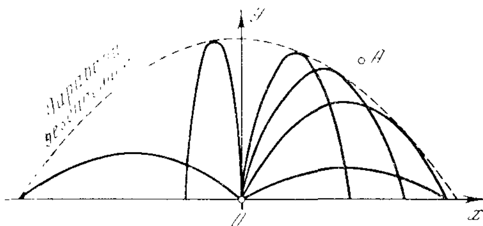


Рис. 9.9.3

время t_2 полета снаряда в 2 раза больше времени t_1 подъема: время подъема снаряда равно времени его падения.

Перейдем теперь от стрельбы по наземным целям к *зенитной* стрельбе. Если при заданном значении начальной скорости v_0 снаряда (эта скорость зависит от орудия) менять угол наклона орудия φ, то мы получим множество разных траекторий снаряда (рис. 3). При этом все навесные траектории, т. е. все параболы, отвечающие значениям φ, заключенным в интервале $45° \leqslant \varphi \leqslant 90°$, как нетрудно видеть, касаются одной и той же линии

$$y = \frac{v_0^2}{2g} - x^2 \frac{g}{2v_0^2} \qquad (9.9.13)$$

(а настильные траектории находятся внутри «колпака», ограниченного линией (13), и самой этой линии не касаются). В самом деле, решая совместно уравнения (9) и (13), находим, что соответствующие линии (параболы) имеют е д и н с т в е н н у ю общую точку:

$$x = \frac{v_0^2}{g}\operatorname{ctg}\varphi, \quad y = \frac{v_0^2}{2g}(1 - \operatorname{ctg}^2\varphi) \qquad (9.9.14)$$

(эта точка расположена в верхней полуплоскости лишь при $\varphi \geqslant 45°$, в силу чего лишь навесные трактории (9) касаются линии (13)). (То, что параболы (9) и (13) в точке (14) к а с а ю т с я, следует уже из того, что общая точка (9) и (13) единственна; это можно подтвердить и прямым вычислением угла наклона касательной к обеим линиям в точке (14); ср. с. 69.) Парабола (13) называется *параболой безопасности*: если цель A (скажем, самолет) расположена вне ограниченной (13) и осью x области, то (зенитный) снаряд не может поразить A ни при каком угле наклона орудия.

Отметим в заключение, что реальные траектории снарядов не являются в точности параболами: в действительности снаряд испытывает сопротивление воздуха, которым мы пренебрегли; к тому же ускорение свободного падения g зависит от высоты над поверхностью Земли; определенные поправки в процесс падения снаряда вносит вращение Земли, да и ньютоновы уравнения движения материальной точки (хотя этот фактор и является менее важным) в применении к снаряду не являются абсолютно точными. Но ведь с подобным положением дела мы встречаемся при л ю б о м применении математики к явлениям реальной жизни: мы всегда пренебрегаем при этом теми или иными факторами, ограничиваясь «огрубленной» картиной, позволяющей использовать математический аппарат, которым мы располагаем. При этом реальная дальность полета снаряда, высота его подъема, полетное время и т. д. зависят от массы снаряда, его формы и плотности воздуха. Так, если при небольшой начальной скорости снаряда (см. отвечающий значению $v_0 = 80$ м/с рис. 2) роль сопротивления воздуха действительно невелика, то при больших v_0 им пренебрегать никак нельзя; например, при $v_0 = 800$ м/с для снаряда калибра (диаметра) 305 мм при угле бросания $\varphi = 55°$ сопротивление воздуха уменьшает дальность полета снаряда от получающегося по формуле (10) значения 61 км до 22,2 км.

Упражнения

9.9.1. Снаряд вылетает из орудия со скоростью 80 м/с. Определить дальность стрельбы и максимальную высоту подъема снаряда, если угол бросания $\varphi = 30°$, 45°, 60°.

9.9.2. Определить наибольшую высоту, на которой снаряд с начальной скоростью $v_0 = 80$ м/с может поразить цель, расположенную на расстоянии 500 м от местонахождения орудия.

§ 10. Движение тел в космическом пространстве

Обратимся теперь к проблемам движения тел (искусственных спутников Земли, космических кораблей) в космическом пространстве. Для того чтобы запущенное с Земли тело стало спутником Земли, т. е. начало вращаться вокруг нее, нужно, чтобы центробежная сила этого тела уравновешивала силу притяжения Земли. Соответствующая скорость u_1 называется *первой космической скоростью* [1]. Для определения величины u_1 служит уравнение

$$m\frac{u_1^2}{R}=mg, \qquad (9.10.1)$$

где g — ускорение свободного падения, m — масса рассматриваемого тела, а R — радиус орбиты (левая часть (1) указывает центробежную силу тела, а правая — силу, с которой тело притягивается к Земле). В случае тела, вращающегося по близкой к Земле орбите, радиус R орбиты близок к радиусу r_0 Земли, поэтому в правой части (1) в качестве силы притяжения может быть взята сила тяжести на поверхности Земли. Из формулы (1) следует

$$u_1=\sqrt{gR}\approx\sqrt{gr_0}\approx 8 \text{ км/с}. \qquad (9.10.2)$$

Для спутника, летящего на расстоянии R от центра Земли, значительно отличающемся от радиуса Земли r_0, надо учесть, что с изменением высоты меняется величина g ускорения свободного падения, которую на поверхности Земли можно обозначить через g_0 (именно это значение $g=g_0$ мы подставляли в формулы (1) и (2)). Действительно, по закону тяготения Ньютона тело, отстоящее от центра Земли на расстояние R, притягивается к Земле с силой $F=\gamma\frac{mM}{R^2}$, где m — масса тела, M — масса Земли. Кроме того, по второму закону Ньютона $F=mg$, где $g=g(R)$ — ускорение свободного падения на расстоянии R от центра Земли. Сравнивая два выражения для F, находим $g=\frac{\gamma M}{R^2}$. Если $r=r_0$, то $g=g_0$, поэтому $g_0=\frac{\gamma M}{r_0^2}$, откуда $\gamma=\frac{g_0r_0^2}{M}$ и, значит,

$$g(R)=g_0\left(\frac{r_0^2}{R^2}\right).$$

Условие (1) равенства центробежной силы и силы тяжести теперь принимает вид

$$m\frac{u_1^2}{R}=mg_0\frac{r_0^2}{R^2},$$

и, значит, скорость u_1 спутника на орбите должна быть равна

$$u_1=\sqrt{\frac{g_0r_0^2}{R}}.$$

Чем больше расстояние R, тем меньше скорость u_1, необходимая для того, чтобы спутник оставался на соответствующей орбите. Однако это вовсе не означает, что легче запустить спутник на орбиту с весьма большим R, чем на орбиту с R, близким к r_0: ведь для вывода спутника на орбиту большого радиуса R надо затратить большую энергию на преодоление силы тяжести на пути от земной поверхности до орбиты. Практическое значение имеет движение по круговой орбите, расположенной над земным экватором на высоте примерно 36 000 км: запущенный на такую высоту спутник будет «висеть» над одной точкой экватора (см. упр. 1).

Рассмотрим теперь следующую по трудности осуществления задачу. Для того чтобы тело могло уйти из области действия земного притяжения, нужно, чтобы его начальная кинетическая энергия была больше разности потенциальной энергии удаленного тела и тела на поверхности Земли. Эту величину мы нашли в § 2 (формула (2.8)). При этом предполагается, что приобретение скорости происходит быстро на участке пути, малом по сравнению с радиусом Земли, так что изменением потенциальной энергии на этом участке пути можно пренебречь. Другими словами, мы считаем, что забрасывающая наше тело реактивная сила, действующая в период горения реактивного топлива (см. § 11), весьма велика и действием силы тяжести за это время можно пренебречь [2].

[1] Мы сразу рассматриваем установившийся процесс движения со скоростью u_1 (а далее — со скоростью u_2 и u_3), игнорируя переходную стадию запуска тела с Земли.

[2] Доказано, что быстро сжигать топливо выгоднее (меньше потребуется топлива), не-

Начальная скорость u_2, которую должно иметь тело для того, чтобы уйти из области земного притяжения, называется *второй космической скоростью.* Найдем ее. По формуле (2.8) начальная энергия, необходимая для достижения расстояния R от центра Земли телом, первоначально находящимся на поверхности Земли на расстоянии r_0 от ее центра, равна $K_0 = mg\frac{r_0}{R}(R - r_0)$. В нашем случае R значительно больше r_0, поэтому $R - r_0 \approx R$, что дает $K_0 \approx mgr_0$. Приравняем этой величине кинетическую энергию ракеты:

$$m\frac{u_2^2}{2} = mgr_0;$$

отсюда

$$u_2 = \sqrt{2gr_0} \approx 11{,}2 \text{ км/с}. \qquad (9.10.3)$$

Из (2) и (3) следует:

$$u_2 = \sqrt{2}\,u_1 \approx 1{,}4u_1, \quad u_2^2 = 2u_1^2.$$

Наконец, начальная скорость u_3, которой должно обладать тело для того, чтобы уйти из поля тяготения Солнца, т. е. за пределы солнечной планетной системы, называется *третьей космической скоростью.* Найдем и ее, пользуясь тем, что скорость v_1 вращения Земли вокруг Солнца известна: $v_1 \approx 30$ км/с.

По закону тяготения Ньютона сила притяжения тела массы m к Солнцу есть $F = \frac{-kM_1m}{r^2}$, где M_1 — масса Солнца ($M_1 \approx 2 \cdot 10^{30}$ кг), r — расстояние от центра Солнца до тела, k — постоянный коэффициент. Потенциальная энергия тела, удаленного на расстояние r от центра Солнца, равна

$$u(r) = -\frac{kM_1}{r}m. \qquad (9.10.4)$$

При этом за нуль принято значение потенциальной энергии на бесконечности (ср. § 2).

Величину потенциальной энергии тела, находящегося на радиусе земной орбиты, легко выразить через скорость движения Земли по орбите. В самом деле, на орбите Земли сила притяжения Земли к Солнцу уравновешивается центробежной силой, т. е.

$$M_2\frac{v_1^2}{r_1} = k\frac{M_1M_2}{r_1^2}, \qquad (9.10.5)$$

где v_1 — скорость вращения Земли вокруг Солнца, r_1 — радиус земной орбиты ($r_1 \approx 150 \cdot 10^6$ км$=1{,}5 \cdot 10^{11}$ м) и M_2 — масса Земли (она, впрочем, выпадает из уравнения (5)). Отсюда

$$kM_1 = v_1^2r_1,$$

и (4) принимает вид

$$u(r_1) = -v_1^2m.$$

Для того чтобы тело, находившееся на расстоянии r_1 от Солнца, ушло из поля тяготения Солнца, нужно, чтобы при таком удалении от Солнца сумма его кинетической и потенциальной энергий была неотрицательна. Последнее условие приводит к неравенству

$$m\frac{v_2^2}{2} + u(r_1) = m\frac{v_2^2}{2} - mv_1^2 \geqslant 0. \qquad (9.10.6)$$

Здесь v_2 — искомая скорость, необходимая для ухода из солнечной системы, v_1 — известная скорость движения Земли.

С подобным положением мы уже встречались при рассмотрении движения тела в поле тяготения Земли: скорости, необходимой для ухода тела из поля земного тяготения, соответствует кинетическая энергия, равная у д в о е н н о й кинетической энергии, отвечающей скорости, обеспечивающей вращение запущенного тела вокруг Земли в качестве (искусственного) ее спутника.

Из соотношения (6) находим минимальную скорость v_2, необходимую для ухода из солнечной системы:

$$v_2 = \sqrt{2}\,v_1 \approx 1{,}4v_1 \approx 42 \text{ км/с}.$$

Итак, для того чтобы уйти из солнечной системы, необходимо на орбите Земли иметь начальную скорость (относительно Солнца) 42 км/с. При этом оказывается, что если скорость тела больше 42 км/с, то тело уйдет из солнечной системы независимо от направления скорости — независимо от того, движется ли тело прямо по радиусу от

жели растягивать процесс горения на все время, необходимое для прохождения пути порядка радиуса Земли. Лишь на том участке пути, где еще велики плотность атмосферы и сопротивление воздуха, невыгодно двигаться с большой скоростью. Однако толщина атмосферы мала по сравнению с радиусом Земли (см. гл. 11), и мы ее учитывать не будем.

Солнца (см. траекторию *1* на рис. 1), или по касательной к орбите Земли (траектории *2*, *3*), или даже в сторону Солнца (но под некоторым углом к направлению на Солнце, чтобы не попасть на его поверхность, — траектория *4*). От направления начальной скорости зависит только форма траектории (см. рис. 1).

Ясно, что для запуска ракеты с Земли наиболее выгодна траектория *2*: сама Земля движется со скоростью 30 км/с, поэтому для получения скорости 42 км/с в этом же направлении достаточно, чтобы относительно Земли ракета двигалась со скоростью $v_2'=12$ км/с. Эту скорость v_2' ракета должна иметь после того, как она выйдет из поля тяготения Земли, т. е. удалится от Земли на расстояние, большое по сравнению с радиусом Земли, но малое по сравнению с радиусом земной орбиты.

Какую же начальную скорость u_3 для этого надо иметь на поверхности Земли (именно эта скорость и называется третьей космической скоростью)? Определим u_3 из соотношения

$$m\frac{u_3^2}{2}=mgr_0+m\frac{(v_2')^2}{2}. \qquad (9.10.7)$$

Здесь первый член справа есть энергия, необходимая для преодоления притяжения Земли, второй член — энергия, которая должна остаться после преодоления притяжения Земли для того, чтобы ракета (или космический корабль массы m) имела (в сумме со скоростью движения Земли) скорость v_2, необходимую для ухода из солнечной системы. Формула (7) дает

$$u_3^2=2gr_0+(v_2')^2=u_2^2+(v_2')^2, \qquad (9.10.8)$$

откуда

$$u_3=\sqrt{u_2^2+(v_2')^2}\approx\sqrt{11{,}2^2+12^2}\approx$$
$$\approx 16{,}4 \text{ км/с}. \qquad (9.10.9)$$

Заметим, что, для того чтобы приблизиться к Солнцу или попасть, например, на Меркурий или Венеру, второй космической скорости недостаточно. Действительно, при наличии этой скорости ракета отделится от Земли и будет двигаться по орбите со скоростью, равной скорости Земли, т. е. 30 км/с.

Хотя потенциальная энергия и уменьшается при приближении к Солнцу, но ракете трудно приблизиться к Солнцу, так как этому мешает центробежная сила ее движения по орбите. Для того чтобы проникнуть в глубь солнечной системы, необходимо уменьшить скорость ракеты относительно Солнца, а это так же сложно, как и

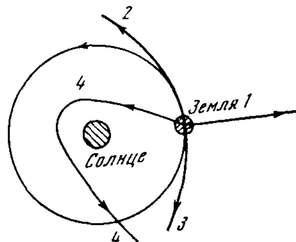


Рис. 9.10.1

увеличение ее. В частности, для попадания в Солнце нужно ракету остановить, т. е. нужно, чтобы она имела скорость 30 км/с относительно Земли (после выхода из поля силы тяжести). Для этого нужно, чтобы на поверхности Земли ракета имела начальную скорость

$$u_4\approx\sqrt{30^2+11{,}2^2}\approx 32 \text{ км/с}. \qquad (9.10.10)$$

Попасть в Солнце труднее, чем уйти от него! Более выгодные варианты можно получить, используя изменение скорости ракеты под влиянием других планет. На этом вопросе мы здесь не остановимся.

Упражнение

9.10.1. Найдите радиус орбиты такой, что время, требующееся для одного оборота вокруг Земли запущенного на эту орбиту спутника, равно 24 ч. (Спутник, запущенный на такую орбиту, в плоскости экватора неподвижно висит над одним пунктом Земли.)

§ 11. Реактивное движение и формула К. Э. Циолковского

Широко применяемый (и достаточно эффективный) способ управления полетом в безвоздушном пространстве, т. е. способ, позволяющий изменять скорость и направление движения, заключается в отбрасывании части массы самого летящего тела, т. е. в применении принципа *реактивного движения*.

Основные закономерности реактивного движения впервые исследовал выдающийся русский ученый и глубокий мыслитель Константин Эдуардович Циолковский (1857—1935), который

указал на возможность использования ракет для освоения космического пространства. От него, через его учеников и последователей — советских ученых и конструкторов идет научная традиция, воплощением которой явились советские искусственные спутники Земли, искусственные планеты и космические корабли с космонавтами на борту.

Выведем основное уравнение прямолинейного движения ракеты. Топливо ракеты — будь то порох или смесь горючего (спирта, бензина) и окислителя (кислорода, азотной кислоты) — обладает определенным запасом Q химической энергии на единицу массы (Q порядка $5\cdot10^6$ Дж/кг для бездымного пороха и $10\cdot10^6$ Дж/кг для бензина с кислородом [1]). При сгорании эта химическая энергия превращается в тепловую энергию продуктов горения. Затем продукты горения вытекают через сопло, при этом их тепловая энергия превращается в кинетическую энергию движения.

Когда реактивный двигатель закреплен на испытательном стенде, продукты горения вытекают с определенной скоростью u_0. При этом кинетическая их энергия на единицу массы составляет определенную долю химической энергии топлива:

$$\frac{u_0^2}{2}=\alpha Q, \qquad (9.11.1)$$

где α — безразмерное число (коэффициент полезного действия (КПД) процессов горения и истечения газов) [2]. В дальнейшем будем считать скорость истечения u_0 известной, заданной величиной. Она составляет около 2 км/с для пороха и около 3 км/с для жидкого топлива. Легко убедиться, что этим величинам соответствуют КПД порядка 50% (т. е. значения $\alpha\approx0{,}5$).

До горения топливо покоилось. Пусть сгорела и вытекла из сопла масса dm топлива. При этом она приобрела количество движения u_0dm. Очевидно, что импульс силы dI, с которой ракета действует на эту массу, равен количеству движения, приобретенному массой [3]:

$$dI=Fdt=u_0dm.$$

По третьему закону Ньютона — закону равенства действия и противодействия — импульс силы, с которой масса dm продукта горения действует на ракету, равен (но противоположен по знаку) импульсу силы, действующей на сами продукты горения. Пусть, например, скорость истечения продуктов горения u_0 направлена в сторону убывания x. Тогда u_0 отрицательно, $u_0=-|u_0|$. Для импульса силы, действующей на ракету, имеем

$$dI_{\text{р}}=F_{\text{р}}dt=-u_0dm=|u_0|\,dm. \qquad (9.11.2)$$

Величина

$$I'=\frac{dI}{dm}=|u_0| \qquad (9.11.3)$$

есть импульс силы, приходящейся на единицу массы, так называемый *единичный импульс*. Эта величина равна скорости истечения газов из покоящейся ракеты.

Проверим размерности в формуле (3). Сила F имеет размерность кг·м/с^2 (ньютон), импульс силы I — это произведение силы на время, поэтому его размерность Н·с (или кг·м/с). Размерность $\frac{dI}{dm}$ есть (кг·м/с)/кг = м/с, т. е. совпадает с размерностью скорости. Для пороховых газов $u_0=2\cdot10^3$ м/с $=2$ км/с, для жидкого топлива $u_0=3$ км/с.

Сила, действующая на ракету, по формуле (2) равна

$$F_{\text{р}}=|u_0|\frac{dm}{dt}.$$

Она пропорциональна количеству продуктов горения, вытекающему из ракеты в единицу времени.

Обратимся теперь к выводу формулы для скорости движения ракеты. Если ракета сама движется с какой-то скоростью u, то скорость истечения газов отлична от u_0 и равна $u+u_0=u-|u_0|$ (напомним, что в покоящейся ракете скорость истечения газов равнялась $-|u_0|$). Очевидно, что такие величины, как р а з н о с т ь с к о р о с т е й по-

[1] Теплотворная способность бензина около $50\cdot10^6$ Дж/кг, однако для сжигания 1 кг бензина нужно израсходовать еще 3,4 кг кислорода. В ракете, летящей в безвоздушном пространстве, кислород надо везти с собой и энергию относить к сумме массы топлива и окислителя.

[2] Если в формуле (1) Q выражено в Дж/кг, то u_0 будет выражено в м/с.

[3] Обозначение dI связано с тем, что рассматривается малая масса dm.

роха до горения и вытекших пороховых газов и сила, с которой на ракету действуют пороховые газы, не зависят от того, движется ракета или покоится (мы далее для определенности говорим о горючем как о порохе).

Обозначим начальную массу ракеты (вместе с порохом) через M_0, а массу вытекших пороховых газов — m. Величина m есть функция времени: $m=m(t)$. Масса ракеты с порохом в момент t равна

$$M=M(t)=M_0-m(t). \qquad (9.11.4)$$

Уравнение движения (второй закон Ньютона) имеет вид

$$M\frac{du}{dt}=F=|u_0|\frac{dm}{dt};$$

его можно переписать так:

$$Mdu=|u_0|\,dm,$$

или, используя (4),

$$(M_0-m)\frac{du}{dm}=|u_0|. \qquad (9.11.5)$$

Возможность сократить dt физически означает, что (при отсутствии других действующих на ракету сил) скорость ракеты зависит только от количества вытекших пороховых газов (при фиксированной величине u_0). К моменту, когда из сопла вытекло данное количество m пороховых газов, ракета приобретает определенную скорость u, не зависящую от того, за какое время произошло это вытекание.

Нетрудно решить дифференциальное уравнение (5). В начальный момент при $m=0$ скорость $u=0$. Поэтому, разделяя в (5) переменные и интегрируя, получаем

$$u=|u_0|\int\limits_0^{m_1}\frac{dm}{M_0-m}=$$

$$=-|u_0|\ln(M_0-m)\Big|_0^{m_1}=$$

$$=|u_0|[-\ln(M_0-m_1)+\ln M_0]=$$

$$=|u_0|\ln\frac{M_0}{M_0-m_1}=|u_0|\ln\frac{M_0}{M},$$

где m_1 — общая масса вытекших к известному моменту пороховых газов, $M=M_0-m_1$ — масса ракеты в этот момент. Таким образом, мы приходим к формуле К. Э. Циолковского

$$u=|u_0|\ln\frac{M_0}{M}. \qquad (9.11.6)$$

Если нас интересует конечная скорость $u_к$ к моменту окончания горения всего топлива, то надо в формулу (6) подставить вместо M величину $M_к$ — конечную массу ракеты после сгорания всего топлива: $M_к=M_0-m_п$, где $m_п$ — полная масса всего топлива. Получим

$$u_к=|u_0|\ln\frac{M_0}{M_к}. \qquad (9.11.7)$$

При помощи этой формулы легко решается и обратная задача: какой должна быть начальная масса M_0 ракеты для того, чтобы данной конечной массе $M_к$ придать определенную скорость $u_к$:

$$\ln\frac{M_0}{M_к}=\frac{u_к}{|u_0|},$$

откуда

$$M_0=M_к e^{\frac{|u_к|}{|u_0|}}. \qquad (9.11.8)$$

При этом в нашем анализе мы всюду игнорировали (переходный) процесс «медленного горения», когда на тело действуют и сила тяжести, и реактивные силы. Реально, разумеется, такое положение дела всегда необходимо имеет место; мы же искусственно разделяли эти две силы, надеясь, что сделанная ошибка не слишком скажется на результатах; при этом учету сил тяготения был посвящен § 10, а изучению реактивных сил — настоящий параграф [4]. Сейчас же мы попытаемся объединить результаты обоих параграфов, использовав формулы (6)—(8) для оценки величины отношения $\frac{M_0}{M_к}$, позволяющей достигнуть 1, 2 и 3-ю космические скорости: u_1, u_2, u_3. В качестве реактивного топлива у нас по-прежнему будет фигурировать порох, обеспечивающий значение $|u_0|=2$ км/с. Формула (8) приводит к следующим результатам: для $u_1=8$ км/с должно быть $\frac{M_0}{M_{к_1}}=$ $=e^4\approx 54$; для $u_2=11{,}2$ км/с должно

[4] Заметим, что с подобным огрублением реального процесса мы встретились также и в § 10, где были искусственно разделены два этапа подъема запущенного в космос тела: на первом этапе мы считали, что ускорение свободного падения $g=g_0=\mathrm{const}$, где g_0 — ускорение свободного падения на поверхности Земли; после этого мы перешли к рассмотрению иного режима движения тела в космосе, характеризующегося уменьшением g при увеличении расстояния от центра Земли.

быть $\frac{M_0}{M_{к_2}}=e^{5,6}\approx 270$; для $u_3=16,4$ км/с должно быть $\frac{M_0}{M_{к_3}}=e^{8,2}\approx 3640$. В случае жидкого топлива, когда $|u_0|=3$ км/с, аналогичные подсчеты дают: $\frac{M_0}{M_{к_1}}\approx 14,5$, $\frac{M_0}{M_{к_2}}\approx 42$, $\frac{M_0}{M_{к_3}}\approx 245$. Из подсчетов видно, что величина $\frac{M_0}{M_к}$ сильно зависит от скорости истечения газов u_0. Для того чтобы оценить трудности задачи по запуску ракеты, следует также иметь в виду, что $M_к$ включает в себя массу баков для топлива и т. п.

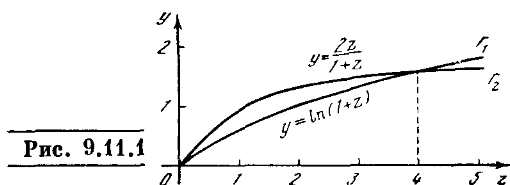


Рис. 9.11.1

Найдем КПД ракеты как целого. Определим эту величину как отношение кинетической энергии ракеты после сгорания топлива $\frac{M_к u_к^2}{2}$ к химической энергии сгоревшего топлива $mQ=(M_0-M_к)Q$. Коэффициент полезного действия (КПД)

$$\eta=\frac{M_к u_к^2}{2Q(M_0-M_к)}. \qquad (9.11.9)$$

Подставив в (9) выражение для $u_к$ из (7) и выразив u_0^2 из (1), окончательно получим

$$\eta=\alpha\frac{M_к}{M_0-M_к}\left(\ln\frac{M_0}{M_к}\right)^2.$$

КПД оказывается произведением «внутреннего» КПД α, характеризующего полноту сгорания топлива и преобразования тепловой энергии в кинетическую энергию газов, и второго множителя, зависящего только от выбора соотношения между массой горючего m и массой полезного груза $M_к$. Обозначим $\frac{m}{M_к}=z$. Тогда $M_0=M_к+m=M_к(1+z)$ и

$$\eta=\alpha\frac{M_к}{m}\left(\ln\frac{M_к+m}{M_к}\right)^2=\frac{\alpha}{z}[\ln(1+z)]^2. \qquad (9.11.10)$$

На первый взгляд может показаться, что при малых z КПД очень велик за счет дроби $\frac{1}{z}$. В действительности при малых z, как мы знаем, имеет место приближенное равенство $\ln(1+z)\approx z$ (см. § 4.9), и потому

$$\eta\approx\frac{\alpha}{z}z^2=\alpha z \quad \text{при } z\ll 1.$$

КПД пропорционален z, и, следовательно, он мал при малом z: при малом z ракета движется медленно, почти всю энергию уносят газы. При очень больших z КПД падает из-за уменьшения полезной массы[5]. Так как окончательная скорость ракеты также зависит только от z, то можно сказать, что КПД ракеты определяется требуемой скоростью. При малой скорости КПД ракеты мал, поэтому невыгодно применение ракет в автомобилях и в других случаях сравнительно медленного движения. При большой скорости энергетический КПД ракеты снова уменьшается, но применение ракет тем не менее оправданно, так как мы не располагаем другими практически удобными способами ускорения тел до больших скоростей.

В заключение определим еще, при какой величине $z=\frac{m}{M_к}$ КПД η достигает наибольшего возможного значения и чему равно это наибольшее значение η_{max} КПД ракеты. В силу (10) задача сводится к отысканию максимума функции $F(z)=\frac{[\ln(1+z)]^2}{z}$, т. е. к решению уравнения

$$F'(z)=\frac{2\ln(1+z)}{(1+z)z}-\frac{[\ln(1+z)]^2}{z^2}=0,$$

или

$$\ln(1+z)=\frac{2z}{1+z}. \qquad (9.11.11)$$

Это уравнение проще всего решить графически, найдя на координатной плоскости точки пересечения графиков Γ_1 и Γ_2 функций

$$f_1=\ln(1+z) \quad \text{и} \quad f_2(z)=\frac{2z}{1+z}.$$

Этим путем (рис. 1) мы получаем $z\approx 4$, откуда $F(z)\approx 0,65$ и $\eta_{max}\approx 0,65\alpha$ (нетрудно проверить, что значению $z\approx 4$ отвечает именно максимум функции $F(z)$).

[5] При большом z величина $[\ln(1+z)]^2$ растет медленнее, чем z. Действительно, обозначая $y=\ln(1+z)$, получим $z=e^y-1$, а функция e^y растет быстрее любой степени y (см. § 6.5).

Упражнение

9.11.1. Докажите, что значению $z \approx 4$ отвечает максимум функции $F(z) = \frac{|\ln(1+z)|^2}{z}$.

§ 12. Масса, центр тяжести и момент инерции стержня

Рассмотрим тонкий стержень. Направим ось x вдоль стержня и обозначим через σ массу, приходящуюся на единицу длины стержня. Таким образом, отрезку длины dx между значениями x и $x+dx$ абсциссы отвечает масса

$$dm = \sigma dx. \tag{9.12.1}$$

Величина σ (в г/см) есть произведение объемной плотности материала d (в г/см^3) и площади S сечения стержня (в см^2): $\sigma = Sd$. Так как стержень может иметь переменные по длине, т. е. зависящие от x сечение и плотность материала, из которого сделан стержень, то σ является функцией координаты: $\sigma = \sigma(x)$. Величину σ естественно назвать *линейной плотностью*, или *плотностью на единицу длины*. Однако так как объемная плотность d нигде дальше в рассмотрение не входит, то сокращенно будем называть σ просто *плотностью*; толщину стержня мы считаем малой и изображаем его просто линией — отрезком оси x. Общая масса стержня, очевидно, равна

$$m = \int_a^b \sigma(x)\,dx, \tag{9.12.2}$$

где a и b — координаты концов стержня.

Пусть наш стержень жестко скреплен с осью x, которую мы здесь представляем себе как жесткий стержень столь малого сечения и массы, что его можно считать не имеющей массы линией; ось x расположена горизонтально, ось y направлена вертикально вверх; сила тяжести действует на стержень, как показано стрелкой, стремясь опустить его (рис. 1).

Представим себе, что ось x — это коромысло весов. На рисунке схематически показана призма, на которую в начале координат опирается ось x. Таким образом, ось x может поворачиваться вокруг оси, перпендикулярной к плоскости чертежа. Найдем, какой груз M, расположенный на оси x слева от начала координат на расстоянии R от него, нужен для того, чтобы уравновесить находящийся справа стержень.

По законам рычага элемент массы dm, находящийся на расстоянии x справа от точки опоры, в том случае уравновешивается элементом массы dM, находящимся слева, если *отношение масс обратно пропорционально отношению расстояний*, т. е. если

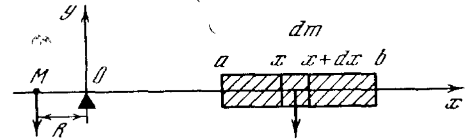


Рис. 9.12.1

$$\frac{dM}{dm} = \frac{x}{R}, \quad \text{или} \quad RdM = xdm. \tag{9.12.3}$$

Элемент массы dm равен, как было выяснено раньше, σdx. Для того чтобы уравновесить весь стержень, нужна такая масса M, что

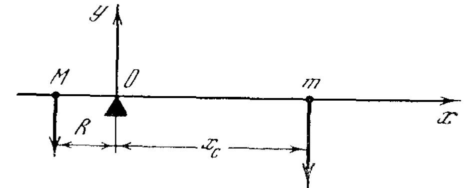


Рис. 9.12.2

$$RM = \int_a^b x\sigma(x)\,dx. \tag{9.12.4}$$

Это равенство получено интегрированием левой и правой частей (3). При этом на правой стороне оси различные элементы массы dm находятся на разных расстояниях x от опоры, поэтому под знак интеграла вошла величина $xdm = x\sigma dx$. (Таким образом, интеграл, от которого зависит уравновешивающая стержень масса M, отличается от интеграла, выражающего массу стержня.) Слева на оси все элементы массы dM (уравновешивающие различные элементы dm стержня) собраны на одном и том же расстоянии R от опоры. R есть постоянная величина, и поэтому $\int RdM = R\int dM = RM$.

Поставим вопрос: если всю массу m стержня сосредоточить в одной точке C,

то на какое расстояние x_C надо удалить эту точку от точки опоры (от начала координат), чтобы уравновесить ту массу M слева, удаленную от точки

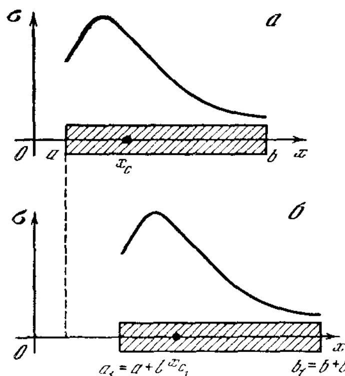


Рис. 9.12.3

опоры на расстояние R, которую уравновешивает стержень (рис. 2)?

Найдем

$$RM = x_C m = \int_a^b x\sigma dx, \qquad (9.12.5)$$

откуда

$$x_C = \frac{1}{m}\int_a^b x\sigma dx = \frac{\int_a^b x\sigma dx}{\int_a^b \sigma dx}. \qquad (9.12.6)$$

Величина x_C есть координата *центра тяжести*, или, как его иначе называют,

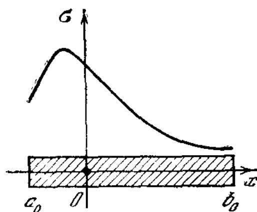


Рис. 9.12.4

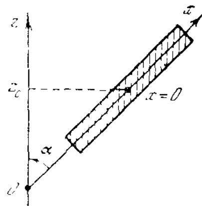


Рис. 9.12.5

центра масс C стержня. Очень важно, что точка C есть действительно вполне определенная точка стержня: если весь стержень как целое переместить вдоль оси x, например, вправо на расстояние l (рис. 3), то и x_C увеличится на ту же величину l, так что для данного стержня точка C с координатой $x=x_C$ находится на определенных расстояниях от концов стержня.

Докажем это.

Рассмотрим стержень, смещенный на расстояние l вправо по сравнению с первоначальным положением (рис. 3, *б*); центр тяжести нового (смещенного) стержня обозначим через C_1. Тогда

$$x_C = \frac{\int_a^b x\sigma(x)\,dx}{\int_a^b \sigma(x)\,dx}, \qquad (9.12.7)$$

$$x_{C_1} = \frac{\int_a^b (x+l)\,\sigma dx}{\int_a^b \sigma dx} = \frac{\int_a^b x\sigma dx + \int_a^b l\sigma dx}{\int_a^b \sigma dx} =$$

$$= \frac{\int_a^b x\sigma dx}{\int_a^b \sigma dx} + \frac{l\int_a^b \sigma dx}{\int_a^b \sigma dx} = x_C + l,$$

что, конечно, полностью совпадает с нашими ожиданиями.

Удобнее всего выбрать систему координат с началом координат, помещенным в центре тяжести стержня (рис. 4). Величины в этой системе координат отметим индексом 0. Разумеется,

$$\int_{a_0}^{b_0} \sigma_0(x)\,dx = m.$$

Так как координата центра тяжести x_{C_0} в такой системе координат равна нулю, то

$$\int_{a_0}^{b_0} x\sigma_0(x)\,dx = 0. \qquad (9.12.8)$$

Другими словами, пожелав уравновесить стержень с точкой опоры в центре тяжести грузом M, расположенным на фиксированном расстоянии R от точки опоры, мы найдем, что уравновешивает наш стержень груз массы $M=0$ (ср. (4) и (8)): если точка опоры совпадает с центром тяжести стержня, то стержень находится в равновесии без всякого дополнительного груза.

Покажем, что при любом положении стержня его **п о т е н ц и а л ь н а я** энергия в поле силы тяжести равна потенциальной энергии всей точечной массы, равной массе стержня и сосредоточенной в центре тяжести стержня.

Потенциальная энергия du элемента стержня массы dm на рис. 5 равна $gzdm$, где z — высота этого элемента стержня, g — ускорение свободного падения. Потенциальная энергия u всего стержня находится интегрированием. В качестве переменной интегрирования выберем длину, отсчитываемую вдоль стержня от его центра тяжести; плотность в точке x обозначаем через $\sigma_0(x)$. Выразим высоту z через x. Как видно из рис. 5, $z(x)=z_C+x\cos\alpha$, где z_C есть высота центра тяжести стержня. Поэтому

$$\int_{a_0}^{b_0} gz\sigma_0(x)dx = g\int_{a_0}^{b_0}(z_C+x\cos\alpha)\,\sigma_0(x)\,dx =$$

$$= gz_C\int_{a_0}^{b_0}\sigma_0(x)\,dx +$$

$$+ g\cos\alpha\int_{a_0}^{b_0}x\sigma_0(x)\,dx = gz_Cm,$$

так как второй интеграл равен нулю в силу (8). Итак, потенциальная энергия зависит только от массы стержня и высоты его центра тяжести, но не зависит от угла, образуемого стержнем с горизонтальным направлением: поворот стержня вокруг его центра тяжести не требует затраты энергии; закрепленный на шарнире в центре тяжести стержень способен свободно (без выделения или затраты энергии) вращаться вокруг точки подвеса.

Перейдем теперь к так называемому *моменту инерции* стержня. Это понятие возникает при рассмотрении вращательного движения стержня. Пусть расположенный в плоскости чертежа стержень вращается вокруг оси, перпендикулярной плоскости чертежа; точку, в которой ось вращения пересечет эту плоскость, мы примем за начало координат. Каждая точка стержня при вращении описывает окружность с центром в начале координат, радиус которой равен абсциссе x данной точки в начальном (горизонтальном) положении стержня (рис. 6). Угловую скорость вращения (в радианах в секунду) обозначим через ω; тогда за время dt стержень повернется на угол $d\varphi=\omega dt$. Длина дуги, которую проходит произвольно выбранная точка с абсциссой x, равна

$$dl = xd\varphi = x\omega dt;$$

следовательно, линейная скорость движения точки стержня, имеющей абсциссу x, равна

$$v(x) = \frac{dl}{dt} = \omega x.$$

Найдем кинетическую энергию вращающегося стержня. Элемент массы dm, удаленный на расстояние x от центра вращения (т. е. от начала

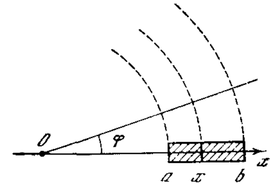


Рис. 9.12.6

координат; здесь речь идет об отрезке стержня длиной dx с координатами концов x и $x+dx$), имеет кинетическую энергию

$$\frac{v^2}{2}dm = \frac{\omega^2x^2}{2}dm = \frac{\omega^2x^2}{2}\sigma(x)\,dx.$$

Следовательно, кинетическая энергия всего стержня

$$E = \frac{\omega^2}{2}\int_a^b x^2\sigma(x)\,dx.$$

Интеграл в последней формуле

$$I = \int_a^b x^2\sigma(x)\,dx \qquad (9.12.9)$$

носит название *момента инерции* стержня относительно оси, проходящей через начало. Таким образом,

$$E = \frac{I\omega^2}{2},$$

т. е. кинетическая энергия вращения выражается через момент инерции и угловую скорость совершенно так же, как кинетическая энергия $E=\frac{mv^2}{2}$ поступательного движения выражается через массу и линейную скорость. Заметим еще, что момент инерции (9) (как и масса m) всегда положителен.

Если мы имеем систему точечных масс m_1, m_2, ..., m_k, расположенных в пунктах M_1,

$M_2, \ldots, M_k$ плоскости, то **моментом инерции** этой системы масс относительно прямой l называется сумма

$$\Sigma m_i d_i^2 = m_1 d_1^2 + m_2 d_2^2 + \ldots + m_k d_k^2 \qquad (9.12.10)$$

произведений массы m_i на квадрат d_i^2 расстояния от M_i до l (здесь $i=1, 2, \ldots, k$). Если же роль системы точечных масс играет тело (пластинка или, как в нашем случае — стержень) с **непрерывным** распределением масс, то момент инерции приближенно равен моменту инерции системы точечных масс, получаемой при разбиении тела на малые части и замене каждой такой части точечной массой, равной массе этой части тела. Ясно, что это определение, где подразумевается, что число частей тела устремляется к бесконечности, а размеры каждой из частей — к нулю, приводит к выражению момента инерции в виде **интеграла** (ср. с формулой (9))[1]. Аналогично этому подобная (10) сумма, где только квадраты d_i^2 расстояний от M_i до l заменены самими этими расстояниями d_i, называется *статическим моментом* системы масс относительно l:

$$\Sigma m_i d_i = m_1 d_1 + m_2 d_2 + \ldots + m_k d_k. \qquad (9.12.10\text{a})$$

(Здесь, однако, расстояниям d_i приходится приписывать определенный знак в соответствии с тем, по какую сторону от прямой l расположена точка M_i.) Ясно, что переход от системы точечных масс к телу и в этом случае приводит к необходимости замены суммы интегралом; так, статический момент $\int_a^b x\sigma(x)dx$ стержня относительно точки O фигурирует в формуле (6) для центра тяжести стержня.

Обратимся к вычислению интеграла I. Момент инерции стержня, центр тяжести которого совпадает с центром вращения (с началом координат) O, обозначим через I_0:

$$I_0 = \int_{a_0}^{b_0} x^2 \sigma_0(x)\, dx. \qquad (9.12.11)$$

Определим теперь момент инерции стержня, центр тяжести C_1 которого удален на расстояние l от начала координат (мы будем считать, что точка C_1 расположена справа от O, так что $x_{C_1}=l$). В этом случае $a=a_0+l$, $b=b_0+l$, $\sigma(x)=\sigma_0(x-l)$,

$$I = \int_a^b x^2\sigma(x)\, dx.$$

Положим $z=x-l$; тогда $x=z+l$, $dx=dz$. При изменении x от a до b величина z изменяется от a_0 до b_0. Поэтому

$$I = \int_{a_0}^{b_0} (z+l)^2\sigma_0(z)\, dz = l^2\int_{a_0}^{b_0}\sigma_0(z)\, dz + 2l\int_{a_0}^{b_0}\sigma_0(z)\, z dz + \int_{a_0}^{b_0}\sigma_0(z)\, z^2 dz. \qquad (9.12.12)$$

Заметим, что $\int_{a_0}^{b_0}\sigma_0(z)\, dz = m$, а второй интеграл справа в (12) согласно (8) равен нулю; наконец, третий интеграл согласно (11) есть I_0.

Таким образом, формула (12) принимает вид

$$I = ml^2 + I_0. \qquad (9.12.13)$$

Величина ml^2 есть, очевидно, момент инерции точечной массы m, находящейся на расстоянии l от оси вращения (от начала координат). Таким образом, *момент инерции стержня относительно произвольной оси, перпендикулярной к стержню, равен сумме момента инерции стержня относительно оси, проходящей через центр тяжести, и момента инерции точечной массы, равной массе стержня, относительно оси, удаленной от этой массы на расстояние, равное расстоянию центра тяжести стержня от оси вращения.*

Наглядно можно представить себе стержень, закрепленный в центре тяжести на шарнире. Тогда вращение оси и шарнира может не сопровождаться вращением самого стержня, т. е. возможно движение, последовательные стадии которого показаны на рис. 7. Кинетическая энергия E' такого движения равна $\frac{mv_{C_1}^2}{2}$, где v_{C_1} — скорость центра тяжести стержня. Но $v_{C_1}=\omega l$, так что $E' = \frac{\omega^2}{2} ml^2$.

---

[1] В случае плоской пластинки или пространственного тела мы при этом приходим к понятию так называемого *кратного* (*двойного* или *тройного*) *интеграла*, распространенного по площади пластинки или по объему тела (в настоящей книге кратные интегралы не рассматриваются).

Движение, которое мы рассматривали раньше (см. рис. 6), отличается от изображенного на рис. 7 тем, что сам стержень также вращается с угловой скоростью ω вокруг своего центра тяжести. Поэтому кинетическая энергия вращения, показанного на рис. 6, оказывается равной сумме энергии изображенного на рис. 7 вращения и энергии вращения вокруг центра тяжести, равной $I_0\frac{\omega^2}{2}$.

Из вывода формулы видно, что подобное простое сложение энергий при сложении двух движений получается только тогда, когда рассматривается движение центра тяжести, ибо только в этом случае равен нулю интеграл (8).

К понятию центра тяжести стержня можно прийти также и по-иному. Из школьного курса физики хорошо известно, что равнодействующей двух параллельных (и одинаково направленных) сил f_1 и f_2, приложенных в точках M_1 и M_2, является сила f_1+f_2, приложенная в точке C, делящей отрезок M_1M_2 в отношении $M_1C : CM_2=f_2 : f_1$ (рис. 8, *а*). Если f_1 и f_2 — силы тяжести масс m_1 и m_2, то точка C (такая, что $M_1C : CM_2=f_2 : f_1=m_2 : m_1$) называется *центром тяжести* рассматриваемых масс; при этом если принять прямую M_1M_2 за ось абсцисс и считать абсциссы точек M_1 и M_2 равными x_1 и x_2, то абсцисса x_C точки C будет равна $\frac{m_1x_1+m_2x_2}{m_1+m_2}$: в самом деле, при этом

$$M_1C : CM_2=\left|x_1-\frac{m_1x_1+m_2x_2}{m_1+m_2}\right| : \left|\frac{m_1x_1+m_2x_2}{m_1+m_2}-x_2\right|=\left|\frac{m_2(x_1-x_2)}{m_1+m_2}\right| : \left|\frac{m_1(x_1-x_2)}{m_1+m_2}\right|=m_2 : m_1.$$

Аналогично, если мы имеем три (расположенные на одной прямой) массы m_1, m_2 и m_3, приложенные в точках M_1 (x_1), M_2 (x_2) и M_3 (x_3) (рис. 8, *б*), то равнодействующей сил тяжести, приложенных к первым двум массам, будет сила, отвечающая массе m_1+m_2 и приложенная в точке $C_{12}\left(\frac{m_1x_1+m_2x_2}{m_1+m_2}\right)$. Равнодействующая же всех трех масс будет приложена в центре тяжести масс m_1+m_2 и m_3, т. е. в точке C с абсциссой

$$x_C=\left[(m_1+m_2)\frac{m_1x_1+m_2x_2}{m_1+m_2}+m_3x_3\right]\Big/\big[(m_1+m_2)+m_3\big]=(m_1x_1+m_2x_2+m_3x_3)/(m_1+m_2+m_3).$$

Совершенно так же доказывается и то, что равнодействующая сил тяжести масс $m_1, m_2, \ldots, m_k$, приложенных в точках M_1 (x_1), M_2 (x_2), ..., M_k (x_k), приложена в точке

$$C\left(\frac{m_1x_1+m_2x_2+\ldots+m_kx_k}{m_1+m_2+\ldots+m_k}\right)=C\left(\frac{\sum_i m_ix_i}{\sum_i m_i}\right) \qquad (9.12.14)$$

(см. упр. 4а).

Аналогично этому устанавливается также, что если k масс $m_1, m_2, \ldots, m_k$ в пространстве

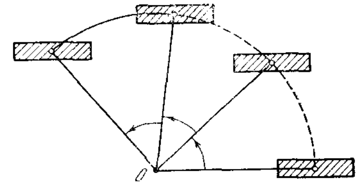


Рис. 9.12.7

приложены в точках M_1 (x_1, y_1, z_1), M_2 (x_2, y_2, z_2), ..., M_k (x_k, y_k, z_k), то совокупность всех отвечающих этим массам сил тяжести может быть заменена одной силой, соответствующей массе $m_1+m_2+\ldots+m_k$ и приложенной в центре тяжести всех этих масс — в точке C (x_C, y_C, z_C) с координатами

$$x_C=\frac{\sum_i m_ix_i}{\sum_i m_i}, \quad y_C=\frac{\sum_i m_iy_i}{\sum_i m_i}, \quad z_C=\frac{\sum_i m_iz_i}{\sum_i m_i} \qquad (9.12.15)$$

(см. упр. 4б).

Обратимся теперь к стержню линейной плотности $\sigma=\sigma(x)$. Если мы разобьем стер-

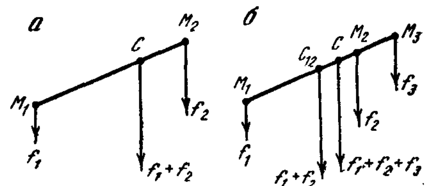


Рис. 9.12.8

жень точками $x_0=a, x_1, x_2, \ldots, x_n=b$ на n малых частей масс $m_i=\sigma(x_i)\,\Delta x_i$, где $\Delta x_i=x_i-x_{i-1}$ и $i=1, 2, \ldots, n$, а затем рассмотрим n точечных масс m_i, отвечающих n частям стержня (предполагаемым однородными, ибо плотность в пределах каждого отрезка Δx_i стержня меняется мало) и сосредоточенных

в конце $x=x_i$ каждой из этих частей, то равнодействующая этих (точечных) масс будет приложена в их центре тяжести $C_1(x_{C_1})$, где $x_{C_1}=\dfrac{\sum\limits_i \sigma(x_i)\,x_i\Delta x_i}{\sum\limits_i \sigma(x_i)\,\Delta x_i}$. Устремляя n к бесконечности, а все Δx_i — к нулю и заменяя сумму интегралом, мы и придем к формуле (6) для центра тяжести стержня.

Упражнения

9.12.1. Найдите момент инерции относительно центра тяжести стержня длины l с равномерно распределенной массой.

9.12.2. Стержень составлен из двух кусков: первый кусок длины l_1 имеет постоянную плотность σ_1, второй кусок длины l_2 имеет также постоянную, но другую плотность σ_2. Найдите положение центра тяжести стержня.

9.12.3. Найдите положение центра тяжести и величину момента инерции относительно центра тяжести для однородного стержня, имеющего форму тонкого треугольника длины L. Выразите эти величины через длину L и массу стержня m. [Указание. Если направить ось x по медиане, а за начало координат принять вершину треугольника, то $\sigma(x)=ax$, где a — постоянная.]

9.12.4. Докажите: а) формулу (14); б) формулы (15).

§ 13*. Центр тяжести нити и пластинки

Выше (см. § 7.10) мы встречались с понятием **центра тяжести** плоской **пластинки** или как угодно искривленной **струны** (нити). Пусть AB, где $A=A(a_1, b_1)$, $B=B(a_2, b_2)$, — (плоская) материальная **нить** произвольной формы и линейной плотности $\sigma=\sigma(x, y)$ (рис. 1); другими словами, мы считаем, что малый участок нити длины $ds=\sqrt{(x')^2+(y')^2}\,dt$, где $x=x(t)$, $y=y(t)$ — параметрические уравнения линии AB, имеет массу $dm=\sigma(x, y)\,ds$, где (x, y) — какая-то точка рассматриваемого участка нити, например его конец. Заменив непрерывное распределение масс, т. е. нашу материальную нить, на систему n точечных масс $dm_i=\sigma(x_i, y_i)\,ds_i$ $(i=1, 2, \ldots, n)$, отвечающих отрезкам ds_i нити с концами $A_{i-1}(x_{i-1}, y_{i-1})$ и $A_i(x_i, y_i)$ $(A_0=A, A_1, A_2, \ldots, A_n=B$ — какие-то точки нити, разбивающие ее на n малых частей), мы придем к выводу, что центр тяжести C_1 этих n точечных масс имеет координаты

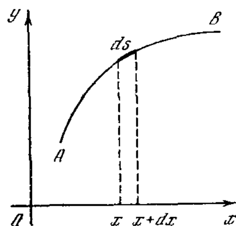


Рис. 9.13.1

$$x_{C_1}=\frac{\sum\limits_i x_i\sigma(x_i, y_i)\,ds_i}{\sum\limits_i \sigma(x_i, y_i)\,ds_i},\quad y_{C_1}=\frac{\sum\limits_i y_i\sigma(x_i, y_i)\,ds_i}{\sum\limits_i \sigma(x_i, y_i)\,ds_i} \tag{9.13.1}$$

(ср. (12.14), (12.15)). Переходя теперь к пределу при $n\to\infty$, $ds_i\to 0$, мы придем к следующему выражению для координат x_C, y_C центра тяжести C нити:

$$x_C=\frac{\int\limits_A^B x\sigma(x, y)\,ds}{\int\limits_A^B \sigma(x, y)\,ds},\quad y_C=\frac{\int\limits_A^B y\sigma(x, y)\,ds}{\int\limits_A^B \sigma(x, y)\,ds}; \tag{9.13.2}$$

здесь запись $\int\limits_A^B$ означает, что соответствующий интеграл распространен по дуге AB. Так, если $x=x(t)$, $y=y(t)$ — параметрические уравнения линии AB и концам A, B отвечают значения t_1, t_2 параметра, то вдоль нити $\sigma=\sigma(x(t), y(t))=\sigma(t)$, общая масса m нити равна

$$m=\int\limits_{t_1}^{t_2}\sigma(t)\sqrt{(x')^2+(y')^2}\,dt$$

и координаты центра тяжести таковы:

$$x_C=\frac{1}{m}\int\limits_{t_1}^{t_2}x(t)\,\sigma(t)\sqrt{(x')^2+(y')^2}\,dt,$$

$$y_C=\frac{1}{m}\int\limits_{t_1}^{t_2}y(t)\,\sigma(t)\sqrt{(x')^2+(y')^2}\,dt. \tag{9.13.3a}$$

Если линия AB задается явным уравнением[1] $y=y(x)$ при $a\leqslant x\leqslant b$, то $\sigma=\sigma(x,\ y(x))=\sigma(x)$, масса $m=$

$$=\int_a^b \sigma(x)\sqrt{1+(y')^2}\,dx \text{ и}$$

$$x_C=\frac{1}{m}\int_a^b x\sigma(x)\sqrt{1+(y')^2}\,dx,$$

$$y_C=\frac{1}{m}\int_a^b y(x)\sigma(x)\sqrt{1+(y')^2}\,dx. \qquad (9.13.3б)$$

В частном случае **однородной** нити постоянной плотности $\sigma=\mathrm{const}$, очевидно, $m=\sigma S$, где $S=\int_a^b ds$ — длина нити; поэтому формулы (2), (3а), (3б) здесь можно переписать так:

$$x_C=\frac{1}{S}\int_A^B x\,ds,\quad y_C=\frac{1}{S}\int_A^B y\,ds, \qquad (9.13.4)$$

или

$$x_C=\frac{1}{S}\int_{t_1}^{t_2} x(t)\sqrt{(x')^2+(y')^2}\,dt,$$

$$y_C=\frac{1}{S}\int_{t_1}^{t_2} y(t)\sqrt{(x')^2+(y')^2}\,dt \qquad (9.13.5а)$$

и

$$x_C=\frac{1}{S}\int_a^b x\sqrt{1+(y')^2}\,dx,$$

$$y_C=\frac{1}{S}\int_a^b y(x)\sqrt{1+(y')^2}\,dx. \qquad (9.13.5б)$$

(Заметьте, что во всех формулах для центров тяжести фигурируют **статические моменты** нити относительно осей Ox и Oy; ср. с. 298.)

Несколько сложнее обстоит дело в случае пластинки D (рис. 2; толщину пластинки мы считаем столь малой, что этой толщиной можно пренебречь). Фактически задача определения центра тяжести пластинки сводится к вычислению некоторых двойных интегралов, распространенных по площади пластинки, которых мы в этой книге не касаемся (и даже не определяем); однако эту трудность можно обойти с помощью следующего приема.

Обозначим через $\rho=\rho(x,y)$ плотность пластинки, отнесенную к единице площади; другими словами, будем считать, что массу малого прямоугольника со сторонами dx и dy, где dx и dy весьма малы (см. рис. 2; стороны прямоугольника мы считаем параллельными осям координат), можно приравнять $\rho(x_0,y_0)\times\times dxdy$, где $dxdy$, разумеется, — площадь прямоугольника, а $\rho=\rho(x_0,\ y_0)$ — плотность в какой-то точке $(x_0,\ y_0)$ прямоугольника, например в его вершине (так как размеры прямоугольника малы, то плотность мало меняется в его пределах). Рассечем теперь нашу пластинку вертикальными прямыми $x=x_0=a$, $x=x_1$, $x=x_2$, ..., $x=x_n=b$ на n вертикальных полос (стержней) толщины $\Delta x_i=x_i-x_{i-1}$ $(i=1,\ 2,\ \ldots,\ n)$; здесь a и b — наименьшее и наибольшее значения абсцисс в пределах пластины D (см. рис. 2). При этом можно считать, что i-я полоска (i-й стержень) имеет переменную (зависящую от y) линейную плотность

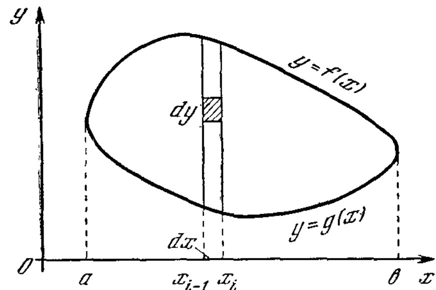


Рис. 9.13.2

$$\sigma_i(y)=\rho(x_i,\ y)\Delta x_i \qquad (9.13.6)$$

и соответственно этому общую массу

$$\Delta m_i=\int_{g(x_i)}^{f(x_i)}[\rho(x_i,\ y)\Delta x_i]\,dy=$$

$$=\left[\int_{g(x_i)}^{f(x_i)}\rho(x_i,\ y)\,dy\right]\Delta x_i, \qquad (9.13.7)$$

где $y=g(x_i)$ и $y=f(x_i)$ — наименьшее и наибольшее значения ординаты y у точек i-го стержня, точнее, $(x_i,\ g(x_i))$ и $(x_i,\ f(x_i))$ — точки пересечения пря-

[1] Здесь мы считаем, что каждому значению x отвечает **единственная** точка нити; если бы это было не так, то мы просто разбили бы всю нить на несколько участков, для каждого из которых выполнено это условие.

мой $x=x_i$ с границей[2] пластины D. (Заметьте, что под интегралом в правой части (7) стоит функция одной переменной y, поскольку значение x_i здесь закреплено!) Эту «распределенную по стержню» массу Δm_i можно заменить точечной массой Δm_i, сосредоточенной в центре тяжести стержня — в точке $C_i\left(x_{C_i}, y_{C_i}\right)$ с координатами

$$x_{C_i}=x_i,$$

$$y_{C_i}=\int\limits_{g(x_i)}^{f(x_i)} y\rho(x_i, y)\,dy\Bigg/\int\limits_{g(x_i)}^{f(x_i)} \rho(x_i, y)\,dy \qquad (9.13.8)$$

(ср. (6) и (12.6)[3]; заметьте, что множитель Δx_i плотности (6) сократится в дроби, выражающей ординату y_{C_i} центра тяжести C_i).

Таким образом, мы приходим к замене действующих на пластинку сил тяжести n силами, отвечающими массам $\Delta m_1, \Delta m_2, \ldots, \Delta m_n$ (ср. (7)) и сосредоточенными в точках $C_1, C_2, \ldots, C_n$ (см. (8)). Далее нам остается только заменить эти n сил тяжести одной силой, приложенной в центре тяжести рассмотренной системы масс, и затем (как обычно!) перейти от сумм к интегралам, условившись, что $n\to\infty$, а все $\Delta x_i\to 0$.

Вывод соответствующих формул не представляет затруднения (см. упр. 2), однако он все же выходит за рамки тем настоящей книги, поскольку здесь приходится встретиться с повторным интегрированием (интегрированием функций, которые сами выражаются интегралами). Поэтому мы не станем загромождать изложение этими формулами, ограничившись случаем однородной пластинки постоянной плотности $\rho=\text{const}$, которую, разумеется, можно принять за единицу: $\rho(x, y)\equiv 1$. В таком случае стержень толщины Δx_i имеет массу

$$\Delta m_i=\int\limits_{g(x_i)}^{f(x_i)} dy\Delta x_i=[f(x_i)-g(x_i)]\,\Delta x_i, \qquad (9.13.9)$$

совпадающую с площадью стержня (произведение его длины $f(x_i)-g(x_i)$ на толщину Δx_i), а центр тяжести C_i стержня расположен, естественно, в его середине, т. е.

$$x_{C_i}=x_i, \quad y_{C_i}=\frac{f(x_i)+g(x_i)}{2}. \qquad (9.13.10)$$

Общая сумма всех масс (9) равна

$$m=\sum_i \Delta m_i=\sum_i [f(x_i)-g(x_i)]\,\Delta x_i.$$

Переходя к пределу при $n\to\infty$, $\Delta x_i\to\to 0$, мы получаем выражение

$$M=\int\limits_a^b [f(x)-g(x)]\,dx,$$

совпадающее, очевидно, с площадью S пластинки D (ср. § 7.4). Координаты же центра тяжести C' масс (9), сосредоточенных в точках (10), имеют вид

$$x_{C'}=\sum_i x_i[f(x_i)-g(x_i)]\,\Delta x_i/S,$$

$$y_{C'}=\sum_i \frac{f(x_i)+g(x_i)}{2}[f(x_i)-g(x_i)]\,\Delta x_i/S$$

(ср. (12.14), (12.15)), или

$$x_{C'}=\sum_i x_i[f(x_i)-g(x_i)]\,\Delta x_i/S,$$

$$y_{C'}=\frac{1}{2}\sum \{[f(x_i)]^2-[g(x_i)]^2\}\,\Delta x_i/S.$$

Тот же предельный переход при $n\to\infty$, $\Delta x\to 0$ приводит к следующим (точным!) формулам для координат x_C, y_C центра тяжести C однородной пластинки D площади S:

$$x_C=\frac{1}{S}\int\limits_a^b x[f(x)-g(x)]\,dx, \qquad (9.13.11)$$

$$y_C=\frac{1}{2S}\int\limits_a^b \{[f(x)]^2-[g(x)]^2\}\,dx,$$

которые мы еще перепишем отдельно для специального случая криволинейной трапеции $D\equiv ABDC$, ограниченной осью абсцисс, прямыми $x=a$ и $x=b$

---

[2] Для простоты мы принимаем, что пластинка D имеет изображенную на рис. 2 «овальную» форму, т. е. ограничена двумя простыми кривыми $y=g(x)$ и $y=f(x)$, взятыми в пределах от $x=a$ до $x=b$.

[3] Формула (12.6) относится к случаю горизонтального стержня; однако уже из определения центра тяжести стержня как такой его точки C, что закрепленный в C стержень находится в равновесии (см. с. 297), вытекает равноправие осей x и y (переходящих одна в другую при повороте стержня вокруг C на 90°); поэтому вторая из формул (8) является следствием формулы (12.6).

и линией $y = f(x)$, т. е. для случая $g(x) \equiv 0$;

$$x_C = \frac{1}{S} \int_a^b x f(x)\, dx,$$
$$y_C = \frac{1}{2S} \int_a^b [f(x)]^2\, dx, \qquad (9.13.12)$$

где, как мы знаем,

$$S = \int_a^b f(x)\, dx.$$

Упражнения

9.13.1. Найдите центр тяжести однородной пластинки, имеющей форму: а) треугольника; б) трапеции; в) полукруга.

9.13.2. Как найти координаты x_C, y_C центра тяжести C пластинки D плотности $\rho = \rho(x, y)$, ограниченной линиями $y = g(x)$ и $y = f(x)$ (в пределах $a \leqslant x \leqslant b$)?

§ 14. Движение тела в среде, противодействующей движению, под действием силы, зависящей только от скорости

Всякое тело испытывает при движении противодействие со стороны той среды, в которой происходит движение. Если сопротивление среды невелико — скажем, при движении с небольшой скоростью в воздухе, — то его можно не принимать во внимание. Однако в ряде случаев такое упрощение недопустимо и с сопротивлением среды приходится считаться.

Экспериментально установлено, что если тело движется в жидкости или газе, скорость движения невелика и размеры тела малы, то сила сопротивления *пропорциональна скорости* движения:

$$F(t) = -kv(t). \qquad (9.14.1)$$

Здесь коэффициент пропорциональности $k > 0$, а знак минус в формуле (1) показывает, что сила сопротивления среды по направлению противоположна скорости движения тела. Число k зависит от свойств среды, оно пропорционально *вязкости*[1] среды. Кроме того, k зависит от формы и размеров тела. Например, для случая шара радиуса R формула (1) принимает вид закона Стокса[2]:

$$F = -6\pi R \eta v(t), \qquad (9.14.2)$$

где η — вязкость среды. Для воздуха $\eta = 1{,}8 \cdot 10^{-4}$, для воды (при 20° С) $\eta = 0{,}01$ г/(см·с).

Рассмотрим задачу о торможении тела. Пусть некоторая сила сообщила телу скорость, а затем (в момент времени $t = t_0$) перестала действовать. Тело продолжает двигаться в вязкой среде, и далее на него действует только сила сопротивления.

В силу второго закона Ньютона

$$m \frac{dv}{dt} = -kv.$$

Разделив обе части на m и обозначив $\frac{k}{m} = \alpha$ $(\alpha > 0)$, получим

$$\frac{dv}{dt} = -\alpha v. \qquad (9.14.3)$$

Уравнение (3), как мы знаем (см., например, гл. 4 и 8), имеет решение

$$v(t) = v_0 e^{-\alpha(t - t_0)}, \qquad (9.14.4)$$

где v_0 $(= v(t_0))$ — значение скорости в момент $t = t_0$. Так как $\alpha > 0$, то при $t > t_0$ показатель степени в (4) отрицателен, $e^{-\alpha(t-t_0)} < 1$ и, следовательно, $v(t) < v_0$, т. е. скорость с течением времени убывает: среда тормозит

---

[1] Вязкость η может быть определена следующим образом. Пусть жидкость (или газ) движется в направлении оси x, причем так, что скорости разных частиц различны и зависят от координаты y. (Ясно, что твердое тело так двигаться не может — оно разрушилось бы.) При этом между соседними слоями жидкости или газа возникает сила трения, пропорциональная разности скоростей соседних слоев, т. е. производной $\frac{dv}{dy}$. Коэффициент пропорциональности в выражении для силы f, приходящейся на 1 см² горизонтальной поверхности, и называется *вязкостью*: $f = \eta \frac{dv}{dy}$ (здесь сила f выражена не в Н, а в Н/см², откуда легко вывести, что размерность коэффициента η есть г/(см · с)).

[2] Стокс Джордж Габриель (1819—1903) — английский математик и физик. Формула (2) справедлива при $\frac{vR\rho}{\eta} < 5$, где ρ — плотность среды. Читатель легко убедится, что величина $\frac{vR\rho}{\eta}$ безразмерна. Она называется *числом Рейнольдса* (Re) по имени английского физика и инженера Осборна Рейнольдса (1842—1912).

движение тела, как и следовало ожидать.

Найдем выражение для пути, пройденного телом. Из (4) следует

$$\frac{dx}{dt}=v_0e^{-\alpha(t-t_0)}, \quad \text{или} \quad dx=v_0e^{-\alpha(t-t_0)}\,dt. \tag{9.14.5}$$

Пусть в начальный момент времени (при $t=t_0$) тело находилось в начале координат: $x(t_0)=0$. Интегрируя (5), получим

$$x(t)=v_0\int_{t_0}^{t}e^{-\alpha(t-t_0)}dt,$$

т. е.

$$x(t)=\frac{v_0}{\alpha}[1-e^{-\alpha(t-t_0)}]. \tag{9.14.6}$$

Пользуясь формулой (6), можно найти весь путь, который тело пройдет после момента t_0 (т. е. после того, как сила перестала действовать) и до момента полной остановки. Для этого заметим, что при очень больших t величина $e^{-\alpha(t-t_0)}$ весьма мала $(e^{-\alpha(t-t_0)})\ll 1)$ и ею вполне можно пренебречь. Поэтому путь, который пройдет тело, равен $\frac{v_0}{\alpha}$.

Рассмотрим теперь падение тела в воздухе (с учетом сопротивления воздуха). Направим ось x вниз к Земле, начало координат поместим на высоте H от Земли (т. е. на Земле $x=H$). Предположим, что движение началось в момент $t=0$ с начальной скоростью v_0. Тогда $x(0)=0$, $v(0)=v_0$. Тело находится под действием двух сил: силы тяжести (она способствует падению тела, т. е. движению его вниз) и силы сопротивления воздуха (она препятствует движению).

Второй закон Ньютона дает

$$m\frac{dv}{dt}=mg-kv. \tag{9.14.7}$$

Разделив все члены (7) на m, получим $\left(\text{так как } \frac{k}{m}=\alpha\right)$

$$\frac{dv}{dt}=g-\alpha v,$$

или

$$\frac{dv}{dt}=\alpha\left(\frac{g}{\alpha}-v\right). \tag{9.14.8}$$

Установим размерность величины $\frac{g}{\alpha}$. Так как $\alpha=\frac{k}{m}$, а $k=-\frac{F}{m}$, то α имеет размерность 1/с. Размерность $\frac{g}{\alpha}$ есть см · с/с^2 = см/с, т. е. $\frac{g}{\alpha}$ имеет размерность скорости[3].

Обозначим $\frac{g}{\alpha}=v_1$. Тогда (8) принимает вид

$$\frac{dv}{dt}=\alpha(v_1-v). \tag{9.14.9}$$

Допустим, что $v_0<v_1$. Тогда правая часть (9) в начале движения положительна, а значит, и левая часть положительна: $\frac{dv}{dt}>0$; поэтому скорость $v(t)$ растет. При этом чем ближе значение v к v_1, тем ближе $\frac{dv}{dt}$ к нулю, тем, следовательно, медленнее растет v. Если бы в некоторый момент времени t_1 оказалось $v(t_1)=v_1$, то v было бы постоянным, так как $v=v_1$ является решением уравнения (9) с начальным условием $v(t_1)=v_1$. Аналогично, если в начале движения $v_0>v_1$, то v приближается к v, но в этом случае v убывает. Поэтому через некоторое время после начала движения тело начинает падать с практически постоянной скоростью $v_1=\frac{g}{\alpha}$ независимо от того, какую скорость оно имело в начале падения. График скорости для случая $v_0=0$ имеет вид, показанный на рис. 1: прямая $v=\frac{g}{\alpha}$ является горизонтальной асимптотой этого графика.

Проведенные рассуждения показывают, что ряд свойств $v(t)$ можно обнаружить, даже не решая уравнение (9). Теперь решим его. Положим $v_1-v=z$. Тогда $\frac{dz}{dt}=-\frac{dv}{dt}$ и (9) перепишется так:

$$\frac{dz}{dt}=-\alpha z.$$

[3] Этот подсчет размерности $\frac{g}{\alpha}$ нужен лишь для проверки правильности наших формул. Размерность $\frac{g}{\alpha}$ видна из формулы (8): так как вычитать можно только величины одной размерности, то отношение $\frac{g}{\alpha}$ должно иметь ту же размерность, что и v.

При $t=0$ должно быть $z=v_1-v_0$. Искомое решение есть $z(t)=(v_1-v_0)e^{-\alpha t}$. Возвращаясь снова к функции $v(t)$, получаем

$$v_1-v(t)=(v_1-v_0)e^{-\alpha t}, \quad \text{или} \quad v(t)=$$
$$=v_1+(v_0-v_1)e^{-\alpha t}. \qquad (9.14.10)$$

Из формулы (10) можно сделать те же выводы, которые мы уже сделали, анализируя «на глазок» исходное уравнение (9). Если $v_0>v_1$, то $v(t)>v_1$, так как $(v_0-v_1)e^{-\alpha t}>0$; если же $v_0<v_1$, то $(v_0-v_1)e^{-\alpha t}<0$, и поэтому $v(t)<v_1$. При этом каково бы ни было v_0, все равно при достаточно больших t величина $e^{-\alpha t}$ мала и практически $v(t)=v_1$.

Из (10) получаем $\left(\text{напомним, что } v=\dfrac{dx}{dt}\right)$

$$\frac{dx}{dt}=v_1+(v_0-v_1)e^{-\alpha t}, \qquad (9.14.11)$$

откуда, учитывая, что $x(0)=0$, имеем

$$x(t)=v_1t+\frac{v_0-v_1}{\alpha}(1-e^{-\alpha t}). \qquad (9.14.12)$$

Если скорость тела и его размеры велики, то сила сопротивления *пропорциональна квадрату скорости*. Из опытов установлено, что в этом случае [4]

$$F=-kS\rho\frac{v^2}{2}, \qquad (9.14.13)$$

где S — площадь сечения тела, ρ — плотность среды. От вязкости среды при этом сила сопротивления практически не зависит. Коэффициент k в этой формуле есть безразмерное число: его величина зависит от формы тела (для хорошо обтекаемых тел величина k может иметь значение порядка 0,03—0,05, а для плохо обтекаемых тел она возрастает до 1,0—1,5). Обозначая $\dfrac{kS\rho}{2}=\varkappa$, получим

$$F=-\varkappa v^2(t). \qquad (9.14.14)$$

Ясно, что $\varkappa$ имеет размерность г/см.

[4] Эта формула справедлива при числах Рейнольдса $\dfrac{Rv\rho}{\eta}>100$. Смысл формулы, приведенной в тексте, заключается в том, что при движении большого тела энергия, затрачиваемая в связи с сопротивлением среды, расходуется не на взаимное трение друг о друга разных слоев жидкости, а на приобретение кинетической энергии жидкостью, вынужденной двигаться, для того чтобы расступиться и пропустить тело. Получите отсюда сами формулу для силы (ср. § 15).

Решим задачу о торможении для силы сопротивления (14). Соответствующее уравнение имеет вид

$$m\frac{dv}{dt}=-\varkappa v^2.$$

Разделив обе части на m и положив $\dfrac{\varkappa}{m}=\beta$, где $\beta>0$, получим

$$\frac{dv}{dt}=-\beta v^2.$$

Отсюда $\dfrac{dv}{v^2}=-\beta dt$. Проинтегрировав, получим $-\dfrac{1}{v}\Big|_{v_0}^{v}=-\beta t\Big|_{t_0}^{t}$, где v_0 — скорость

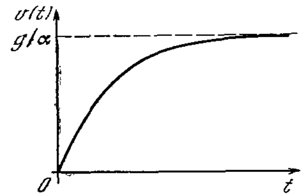


Рис. 9.14.1

тела в момент времени $t=t_0$. Поэтому $-\dfrac{1}{v}+\dfrac{1}{v_0}=-\beta(t-t_0)$, откуда

$$v=\frac{v_0}{1+\beta v_0(t-t_0)}. \qquad (9.14.15)$$

Поскольку $\beta=\dfrac{\varkappa}{m}$, то β имеет размерность 1/см. $\left(\text{Заметьте, что при } t-t_0\gg\dfrac{1}{\beta v_0}\text{ скорость практически не зависит от начальной: } v\approx\dfrac{1}{\beta(t-t_0)}.\right)$

Найдем формулу для пути. Из (15) следует

$$dx=\frac{v_0}{1+\beta v_0(t-t_0)}dt,$$

откуда

$$x(t)=x(t_0)+\int_{t_0}^{t}\frac{v_0}{1+\beta v_0(t-t_0)}dt. \qquad (9.14.16)$$

Если тело начинает движение из начала координат (т. е. если $x(t_0)=0$), то (16) принимает вид

$$x(t)=\frac{1}{\beta}\ln[1+\beta v_0(t-t_0)]. \qquad (9.14.17)$$

Легко убедиться, что формуле (17) соответствует экспоненциаль-

ная зависимость скорости от пройденного пути: $v=v_0e^{-\beta x}$. Если теперь мы захотим найти весь путь, который пройдет тело после того, как сила, сообщившая ему скорость, перестала действовать, то обнаружим, что этот путь (формула (17)) тем больше, чем больше время. (Заметьте, что в силу (17) $x \to \infty$ при $t \to \infty$.) На самом деле это не так. Дело в том, что, когда скорость тела станет малой, соотношение (14) перестанет быть верным. В этом случае надо перейти к формуле (1) и соответственно для подсчета пути — к формуле (6).

Рассмотрим задачу о падении тела в воздухе в случае, когда сопротивление воздуха пропорционально квадрату скорости. Пусть тело падает из начала координат, имея начальную скорость v_0. Совершенно аналогично случаю, когда сопротивление пропорционально первой степени скорости, получаем уравнение

$$\frac{dv}{dt}=g-\beta v^2, \quad \text{или} \quad \frac{dv}{dt}=\beta\left(\frac{g}{\beta}-v^2\right). \qquad (9.14.18)$$

Нетрудно установить, что $\sqrt{\frac{g}{\beta}}$ имеет размерность скорости; обозначим $\sqrt{\frac{g}{\beta}}=v_1$, $\frac{g}{\beta}=v_1^2$. Тогда (18) принимает вид

$$\frac{dv}{dt}=\beta(v_1^2-v^2). \qquad (9.14.19)$$

Точное решение (19) дано в ответах к упр. 1. Рассмотрим общие свойства решения. Рассуждениями, совершенно аналогичными тем, которые были проведены для уравнения (9), показываем, что в этом случае должна установиться скорость движения $v_1=\sqrt{\frac{g}{\beta}}$. Покажем, что через достаточно большое время после начала падения можно будет утверждать

$$v-v_1=Ce^{-2\beta v_1 t}, \qquad (9.14.20)$$

где C — постоянная величина. Уравнение (19) перепишем так:

$$\frac{dv}{dt}=\beta(v_1+v)(v_1-v). \qquad (9.14.21)$$

При больших t имеем $v \approx v_1$, поэтому в (21) заменим v_1+v на $2v_1$. Если заменить v на v_1 в разности v_1-v, получим $\frac{dv}{dt}=0$, откуда $v=\text{const}=v_1$. Так как нас интересует именно малое различие между v и v_1 (закон приближения v к v_1), то пренебречь разностью $v-v_1$ нельзя. Заменяем (21) на

$$\frac{dv}{dt}=2\beta v_1(v_1-v). \qquad (9.14.22)$$

Положим $v_1-v=z$, $\frac{dz}{dt}=-\frac{dv}{dt}$. Теперь (22) примет вид

$$\frac{dz}{dt}=-2\beta v_1 z. \qquad (9.14.23)$$

Решение (23) таково:

$$z=Ce^{-2\beta v_1 t}, \qquad (9.14.24)$$

что совпадает с (20).

Значение C в формуле (24) нельзя определять из начального условия $v(0)=v_0$ (т. е. $z(0)=v_1-v_0$), потому что уравнение (22) справедливо только при достаточно больших t (вблизи $t=0$ заменять $v+v_1$ на $2v_1$ нельзя).

Заметим еще, что формула $F(t)=-\varkappa v^2(t)$ справедлива лишь для случая, когда $v>0$. Действительно, если $v<0$, то должно быть $F(t)=\varkappa v^2(t)$, так как направление силы сопротивления противоположно направлению скорости и, следовательно, эта сила положительна, если скорость отрицательна. Оба случая ($v>0$ и $v<0$) охватывает формула

$$F(t)=-\varkappa v(t)\,|v(t)|.$$

Упражнения

9.14.1. Найдите выражение для скорости в зависимости от времени из уравнения $\frac{dv}{dt}=\beta(v_1^2-v^2)$ при начальном условии $v(0)=0$. Выведите из полученной формулы для v, что происходит установление скорости, равной $v_1=\sqrt{\frac{g}{\beta}}$. Покажите, что для формулы (24) (или (20)) $C=\frac{2v_1(v_0-v_1)}{v_1+v_0}$.

9.14.2. В задаче о падении тела (сила сопротивления пропорциональна скорости) учтите, что на тело действует выталкивающая сила по закону Архимеда.

9.14.3. Применив результат предыдущей задачи к шару и учитывая, что для шара $k=6\pi R\eta$, где R — радиус шара, η — вязкость среды, покажите, что при больших t устанавливается скорость падения шара $v=$

$=\frac{2R^2g(\rho-\rho')}{9\eta}$ (здесь ρ — плотность материала, из которого сделан шар, ρ' — плотность среды).

§ 15*. Движение тел в жидкостях и газах

Остановимся немного подробнее на вопросе о физической природе и закономерностях сил, действующих на тело, движущееся в сплошной среде — в воздухе или в воде. В связи с огромной технической важностью вопроса, в особенности для авиации, он изучен чрезвычайно подробно и составляет главный предмет самостоятельной науки — *гидроаэродинамики*. Частные случаи, которые были рассмотрены в § 14, соответствуют крайнему упрощению и схематизации истинных закономерностей.

Более точным, чем предположения § 14, является утверждение, что силу сопротивления среды, действующую на данное тело, можно записать в виде

$$F=-\frac{kS\rho v^2}{2}, \tag{9.15.1}$$

где S — сечение тела (в см^2), ρ — плотность (в г/см^3) и k — безразмерная величина (ср. формулу (14.14)). При этом в общем случае, если не ограничиваться медленным движением, коэффициент сопротивления k есть функция двух безразмерных величин: *числа Рейнольдса* Re (см. сноску 2 в § 14) и так называемого *числа Маха* [1] — отношения скорости v тела относительно среды к скорости c звука в среде: $\mathrm{Ma}=\frac{v}{c}$. Итак,

$$k=k(\mathrm{Re},\ \mathrm{Ma}). \tag{9.15.2}$$

В частном случае движения тела со скоростью, малой по сравнению со скоростью звука, т. е. при $\mathrm{Ma}\ll 1$, коэффициент сопротивления зависит только от числа Рейнольдса:

$$k(\mathrm{Re},\ \mathrm{Ma}=0)=k(\mathrm{Re}). \tag{9.15.3}$$

Вид функции (3) зависит от формы тела, от его ориентации относительно направления скорости и, наконец, от того, какая именно величина выбрана в качестве характерного размера в определении числа Рейнольдса. Только в самом простом случае — в случае шара — все эти вопросы отпадают.

При малых числах Re, т. е. при $\mathrm{Re}<1$, асимптотика функции (3) такова:

$$k=\frac{\mathrm{const}}{\mathrm{Re}}+\ldots, \tag{9.15.4}$$

где опущены слагаемые, малые по сравнению с выписанным.

Подставляя в (1) выражение для коэффициента k, получаемое при игнорировании невыписанных членов (4), и учитывая, что, по определению, $\mathrm{Re}=\frac{vR\rho}{\eta}$, получим

$$F=-\frac{\mathrm{const}\cdot\eta}{Rv\rho}\frac{\rho v^2}{2}S=\mathrm{const}\cdot\eta Rv, \tag{9.15.5}$$

где, разумеется, запись const в правой и средней частях (5) означает разные числа.

Таким образом, мы снова приходим к закону Стокса (ср. (14.2); характерно, что в этом приближении плотность жидкости не влияет на силу). При этом последнее преобразование в (5) основано на том, что площадь S сечения тела пропорциональна квадрату его линейного размера R (мы рассматриваем геометрически подобные тела); именно поэтому в выражении силы остается первая степень R. Однако надо помнить, что при любом конечном значении числа Re формула (4) является лишь приближенной: она содержит поправки, исчезающе малые по сравнению с выписанным членом лишь при очень малых Re.

Во втором предельном случае при $\mathrm{Re}\to\infty$ ситуация оказывается несколько более сложной. Здесь можно считать k постоянным, как это было сделано выше; хотя предположение о постоянстве k и является довольно грубым, но в книге для начинающих оно, по нашему мнению, допустимо. Во многих учебниках механики и внешней баллистики уравнения движения интегрируются именно в предположении $k=\mathrm{const}$, $F=\mathrm{const}\cdot v^2$. Однако специалист по аэродинамике подчеркнет, что k хоть и слабо, но все же зависит от числа Re даже при больших Re. Так, например, для шара при $\mathrm{Re}=100$ $k\approx 1{,}2$, а при $\mathrm{Re}=2\cdot10^5$ $k\approx 0{,}4$. Затем в сравнительно узкой области $2\cdot10^5<\mathrm{Re}<5\cdot10^5$ сопротивление падает почти втрое, а далее вплоть почти до

[1] Мах Эрнст (1838—1916) — известный австрийский физик, педагог и философ (как с философом с Махом полемизировал Ленин).

$Re = 10^8$ можно считать, что $k = 0{,}12$. В целом при изменении Re на шесть порядков $(10^8 : 10^2 = 10^6)$ k изменилось на один порядок (от 1,2 до 0,12), т. е. в среднем по большому интервалу значений Re получается интерполяция: $k \sim Re^{-1/6}$. При рассмотрении элементарных («прикидочных») расчетов вполне допустима замена слабоменяющейся функции $Re^{-1/6}$ на константу (при условии, что скорость меняется в не слишком широком интервале).

Остановимся теперь на физической картине движения жидкости, обтекающей тело, и на происхождении сил, действующих на него. Нам удобнее перейти в систему координат, связанную с телом, т. е. представить себе покоящееся тело в потоке жидкости.

До встречи с телом, т. е. далеко от него, в направлении, противоположном скорости жидкости, вся жидкость движется с одинаковой скоростью (как по величине, так и по направлению). Вблизи тела движение жидкости меняется. Ясно, что составляющая скорости, перпендикулярная поверхности тела, равна нулю: жидкость не втекает в сплошную стенку тела и не вытекает из тела. Но при наличии вязкости можно утверждать, что на поверхности тела обращается в нуль и составляющая скорости, касательная к поверхности тела: в этом случае скорость жидкости на поверхности тела будет нулевой.

На поверхность тела действует давление, направленное нормально к поверхности, а также сила вязкости, касательная к поверхности и направленная в сторону, в которую движется жидкость вблизи поверхности на малом расстоянии от нее. Сила вязкости равна $\eta \left.\frac{\partial v_t}{\partial n}\right|_S$, где v_t — касательная составляющая скорости, а $\frac{\partial}{\partial n}$ — символ производной в направлении, перпендикулярном к поверхности (в направлении *нормали* к поверхности; на поверхности S тела $v_t = 0$, но производная $\frac{\partial v_t}{\partial n}$ отлична от нуля).

Если известно, как движется жидкость в каждой точке вблизи тела, то рассчитать давление и силу вязкости нетрудно. Но именно расчет движения очень труден и в общем случае нам пока недоступен: до настоящего времени он проведен до конца лишь для малого числа особо простых случаев.

При медленном движении главную роль играют силы вязкости. Распределение скоростей в различных жидкостях и газах, обладающих разной вязкостью, будет подобно при обтекании геометрически подобных тел разных размеров; поэтому производная $\frac{\partial v_t}{\partial n}$ в данной точке тела пропорциональна полной скорости v, деленной на размер R тела. Значит, вязкая сила на единицу поверхности имеет порядок $F_1 = \eta \frac{v}{R}$. Полная же сила $F = SF_1$, но площадь S поверхности геометрически подобных тел пропорциональна квадрату характерного линейного размера: $S = \mathrm{const} \cdot R^2$. Таким образом,

$$F = \mathrm{const} \cdot \eta v R. \qquad (9.15.6)$$

Можно показать, что в случае медленных движений тел учет распределения давления по поверхности меняет только константу. Закон Стокса (14.2) для шара радиуса R, где коэффициент $\mathrm{const} = 6\pi$ (в качестве характерного размера R здесь принят радиус шара), является частным случаем общей формулы (6). Величина 6π в формуле (14.2) найдена расчетом; она хорошо согласуется с опытом. Однако для произвольного несимметричного тела расчет становится весьма трудным даже в предельном случае $Re \to 0$.

Сложнее обстоит дело при обтекании большого тела потоком большой скорости, т. е. при больших числах Рейнольдса. Естественно в этом случае искать движение жидкости с учетом ее инерции; изменение величины и направления движения при этом зависит от неравномерности давления.

При торможении струи возникает давление порядка ρv^2. Так как силы вязкости пропорциональны первой степени скорости, то естественно пренебречь ими при быстром движении, при больших числах Рейнольдса. Казалось бы, цель достигнута, получена формула вида $F_1 = \rho v^2$, $F = \mathrm{const} \cdot \rho v^2 S$ с постоянным коэффициентом сопротивления.

Однако действительность оказалась сложнее! Еще в XVIII в. французский математик и механик Ж. Л. Д'Аламбер показал, что для жидкости, пол-

ностью лишенной вязкости, существует такое решение уравнений движения, при котором давление в разных точках поверхности тела различно, но равнодействующая сила давления тождественно р а в н а н у л ю. В действительности такое течение не осуществляется: даже очень малая вязкость существенно перестраивает поток! Возникает *турбулентность*: несмотря на то что набегающий поток не зависит от времени, вблизи и позади тела возникает хаотически меняющееся (как языки пламени у костра) движение жидкости.

Повышенное давление на передней поверхности тела (на которую набегает поток) не компенсируется давлением позади. Перестройка потока, вызванная влиянием вязкости, оказывается существенной, приводит к определенному не равному нулю сопротивлению. Как уже отмечалось, численное значение этого сопротивления очень слабо зависит от числа Рейнольдса, т. е. слабо зависит от вязкости.

В системе координат, в которой движется тело, а жидкость первоначально (до прихода тела в данную область) покоится, можно грубо считать, что столб жидкости, находящийся на пути, который тело проходит за время t (его масса равна $M_1 = \rho Svt$), при прохождении тела расступается и приобретает при этом скорость порядка v. Следовательно, его энергия равна $E = \frac{M_1 v^2}{2} = \frac{t\rho S v^3}{2}$. Но она заимствована у тела: такую работу должно было произвести тело. Работа, производимая телом за время t, равна силе, умноженной на путь: $A = Fvt$. Приравняв затем A и E, мы из равенства $Fvt = \frac{S\rho v^3 t}{2}$ и получим формулу

$$F = \frac{S\rho v^2}{2} \qquad (9.15.7)$$

для силы, которая верно передает порядок этой величины. Через большое время после прохождения тела жидкость успокаивается, но ее энергия превращается в тепло, а не возвращается телу.

Обтекание тела жидкостью существенно зависит от его формы. Найдены формы, обладающие минимальным сопротивлением, с коэффициентом сопротивления, в десятки раз меньшим, чем у шара или диска, поставленного поперек движения.

Общая идея заключается в том, что жидкость, расступившаяся при вхождении в нее тела, плавно смыкается сзади тела. В этом случае вклад разности давления уменьшается до 10—15% от всей силы сопротивления. Даже при больших числах Рейнольдса бóльшую часть силы, непосредственно действующей на поверхность тела (85—90%), составляет сила вязкости. Однако отсюда вовсе не следует, что общая сила сопротивления и коэффициент k пропорциональны вязкости для хорошо обтекаемых тел при большом значении числа Рейнольдса!

Казалось бы, вязкость входит в формулы для k (и F) в первой степени. Однако вязкая сила пропорциональна $\eta \frac{\partial v_t}{\partial n}$, и оказывается, что при прочих равных условиях (данный размер тела, данная скорость жидкости на бесконечности) при уменьшении η уменьшается переходный слой, в котором меняется скорость, т. е. увеличивается $\frac{\partial v_t}{\partial n}$. Коэффициент сопротивления и для хорошо обтекаемых тел при большом числе Рейнольдса пропорционален вязкости в очень малой степени: он меняется примерно как $\eta^{1/6}$. (Некоторые авторы предпочитают формулы вида $k = (a + b \ln \mathrm{Re})^{-1}$.)

Такова ситуация и в самом крайнем случае, при обтекании тонкой пластинки, расположенной вдоль потока. Казалось бы, в пределе при малой вязкости жидкость может скользить вдоль пластинки — давление никакой составляющей силы не дает, а сила трения пропорциональна вязкости и мала. В действительности же возникает турбулентное движение (при $\mathrm{Re} > 10^6$)[2] и сопротивление увеличивается. Если форма тела или его расположение относительно потока несимметрично, например пластинка наклонена, то кроме силы, направленной параллельно скорости движения жидкости, появляется перпендикулярная составляющая — подъемная сила, которая может более чем в 20 раз превосходить силу сопротивления. На этом основан полет планера и самолета.

---

[2] При подсчете числа Рейнольдса здесь вместо радиуса R шара принято подставлять длину L пластинки: $\mathrm{Re} = \frac{\rho v L}{\eta}$.

Глава 10

КОЛЕБАНИЯ

1. Движение под действием упругой силы

Рассмотрим случай, когда сила, действующая на тело, зависит только от его положения: $F=F(x)$. Выше (см. § 9.2) мы подробно рассматривали работу такой силы и выяснили, что в этом случае система имеет определенную потенциальную энергию $u(x)$, с которой сила связана соотношением

$$F(x)=-\frac{du(x)}{dx}.$$

Обратимся к задаче о движении тела под действием такой силы. Основное уравнение движения имеет вид

$$m\frac{dv}{dt}=F(x). \qquad (10.1.1)$$

Так как в уравнение входит производная по времени, а сила задана как функция координаты x, то непосредственно это уравнение решить нельзя. Естественно для решения задачи искать интересующие нас величины как функции координаты x. Будем искать, в частности, зависимость скорости от координаты, т. е. функцию $v(x)$. Производную по времени $\frac{dv}{dt}$ мы представим как производную сложной функции, так как сама координата x зависит от времени:

$$\frac{dv[x(t)]}{dt}=\frac{dv[x(t)]}{dx}\frac{dx(t)}{dt}.$$

Но $\frac{dx}{dt}$ есть не что иное, как скорость $v(x)$. Таким образом, $\frac{dv}{dt}=v\frac{dv}{dx}$, откуда

$$m\frac{dv}{dt}=mv\frac{dv}{dx}=\frac{d}{dx}\left(\frac{mv^2}{2}\right). \qquad (10.1.1a)$$

Подставляя (1а) в уравнение движения (1), получим

$$\frac{d}{dx}\left(\frac{mv^2}{2}\right)=F(x). \qquad (10.1.2)$$

Интегрируя, находим

$$\int_{v_0}^{v_1}\frac{d}{dx}\left(\frac{mv^2}{2}\right)dx=\int_{x_0}^{x_1}F(x)\,dx,$$

или

$$\frac{mv_1^2}{2}-\frac{mv_0^2}{2}=\int_{x_0}^{x_1}F(x)\,dx.$$

Физический смысл последнего равенства совершенно ясен: *изменение кинетической энергии равно работе силы.*

Используя потенциальную энергию $u=u(x)$, запишем (2) так:

$$\frac{d}{dx}\left(\frac{mv^2}{2}\right)=-\frac{du}{dx},$$

или

$$\frac{d}{dx}\left[\frac{mv^2}{2}+u(x)\right]=0.$$

Если производная какого-нибудь выражения тождественно равна нулю, то само это выражение не меняется. Поэтому

$$\frac{mv^2}{2}+u(x)=\text{const}. \qquad (10.1.3)$$

В таком виде формула (3) выражает закон сохранения энергии: при движении тела под действием силы, зависящей от координаты, остается постоянной сумма кинетической энергии тела $\frac{mv^2}{2}$ и его потенциальной энергии $u(x)$.

Эта сумма называется *полной энергией* тела.

Здесь (как и в § 9.14) мы привели довольно длинные преобразования для того, чтобы показать, что закон сохранения энергии (применительно к механике) есть следствие закона Ньютона. Также следствием закона Ньютона является тот факт, что кинетическая энергия тела есть именно $\frac{mv^2}{2}$, а не какая-нибудь другая функция скорости тела. При этом значение закона сохранения энергии, разумеется, не ограничивается рассмотренными в этой книге простейшими примерами: этот закон сохраняет силу (и значение) в широком классе явлений — даже тех, физическая сущность которых до сих пор остается неясной.

Как решать дальше задачу о движении тела? По значениям скорости v_0 и координаты тела x_0 в начальный момент времени находим полную энергию тела E — величину, остающуюся постоянной на протяжении всего движения: $E=\frac{mv_0^2}{2}+u(x_0)$. При помощи формулы (3), зная E, находим скорость тела в зависимости от x:

$$v(x)=\sqrt{\frac{2}{m}[E-u(x)]}. \qquad (10.1.4)$$

Остается найти связь между x и t. Из (4) получаем

$$\frac{dx}{dt}=\sqrt{\frac{2}{m}[E-u(x)]};$$

отсюда

$$\frac{dx}{\sqrt{\frac{2}{m}[E-u(x)]}}=dt,$$

$$t=t_0+\int_{x_0}^{x}\frac{dx}{\sqrt{\frac{2}{m}[E-u(x)]}}. \qquad (10.1.5)$$

Таким образом, время t выражено как функция координаты x: $t=t(x)$, причем функция эта задана интегралом. Решая уравнение (5) относительно x, можно найти зависимость $x=x(t)$. Однако так как зависимость (4) v от x выражается посредством корня квадратного, то даже простое выражение $u(x)$ часто приводит к весьма сложным выражениям для $t(x)$.

Для того чтобы получить общее представление о характере движения, очень полезно нарисовать кривую $u=u(x)$. Пусть, например, график функции $u(x)$ имеет вид, изображенный на рис. 1. Если провести на этом же рисунке горизонтальную прямую на высоте E (прямую $u=E$), то получится весьма наглядная картина. Здесь интуитивно ясно, что тело будет перемещаться между двумя крайними своими положениями, движение будет иметь характер *колебаний*; скорость будет пропорциональна корню квадратному из разности $E-u(x)$. Так, например, при $x=x_A$ скорость пропорциональна корню квадратному из длины отрезка AB (см. рис. 1). Правее точки C и левее точки D находится область, где $E<u(x)$, т. е. область, в которую тело в полном соответствии со сказанным выше при заданной полной его энергии E проникнуть не может — не хватит энергии (это отражается в появлении в формуле (5) корня квадратного из отрицательного числа). В области же, где $E>u(x)$, для $v(x)$ мы получаем два возможных значения в соответствии с двумя знаками корня:

$$v=\pm\sqrt{\frac{2}{m}[E-u(x)]}.$$

В начальный момент как величина, так и знак скорости $v=v_0$ определяются начальными условиями. Дальше движение происходит в направлении, заданном знаком начальной скорости v_0. Очевидно, что при $v\neq 0$ знак скорости не может внезапно измениться. Так,

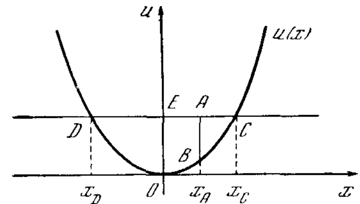


Рис. 10.1.1

если в начальный момент тело находится в точке $x=x_0$ (расположенной где-то между x_D и x_C) и $v_0>0$, то тело дойдет до крайней допустимой точки x_C.

В точке x_C, где скорость тела обращается в нуль, произойдет переход от формулы $v=\sqrt{\frac{2}{m}[E-u(x)]}$ к формуле $v=-\sqrt{\frac{2}{m}[E-u(x)]}$. Так как в этой точке $v=0$, то изменение знака совершается без скачка (разрыва) скорости. Аналогичная картина будет в точке $x=x_D$. Таким образом, в случае, изображенном на рис. 1, движение тела, как мы и ожидаем, будет представлять собой колебания между двумя крайними положениями x_C и x_D.

Рассмотрим другой пример. Пусть потенциальная энергия тела дается функцией $u=ax$, где $a>0$. Найдем закон движения тела.

Обозначим значения координаты x и скорости v в начальный момент времени t_0 через $x=x_0$ и $v=v_0$. Тогда полная энергия $E=\frac{mv_0^2}{2}+ax_0$. Пользуясь (4), получаем

$$v(x) = \sqrt{\frac{2}{m}\left(\frac{mv_0^2}{2} + ax_0 - ax\right)} = \sqrt{v_0^2 + \gamma(x_0 - x)},$$

где $\gamma = \frac{2a}{m}$. Зная $v(x)$, находим время

$$t = t_0 + \int\limits_{x_0}^{x} \frac{dx}{\sqrt{v_0^2 + \gamma(x_0 - x)}}.$$

Поскольку под интегралом стоит выражение $-\frac{1}{\gamma}u^{-1/2}du$, где $u = v_0^2 + \gamma(x_0 - x)$, то

$$t = t_0 - \frac{2}{\gamma}\sqrt{v_0^2 + \gamma(x_0 - x)}\Big|_{x_0}^{x} = t_0 - \frac{2}{\gamma}\left[\sqrt{v_0^2 + \gamma(x_0 - x)} - v_0\right].$$

Отсюда находим x:

$$\frac{\gamma}{2}(t - t_0) = -\sqrt{v_0^2 + \gamma(x_0 - x)} + v_0.$$

Перенося v_0 влево, возводя в квадрат и сокращая на γ, найдем

$$x = -\frac{\gamma}{4}(t - t_0)^2 + v_0(t - t_0) + x_0. \qquad (10.1.6)$$

Вычислив $\frac{d^2x}{dt^2}$, убеждаемся, что движение по закону (6) — равнозамедленное движение. Этого и следовало

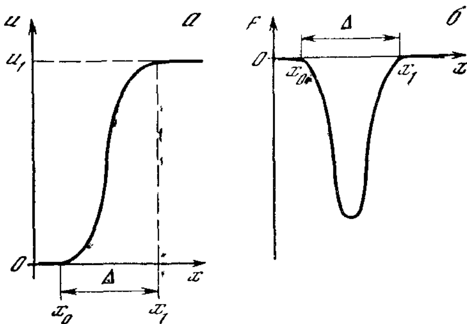


Рис. 10.1.2

ожидать, так как $F = -\frac{du}{dx} = -a$, т. е. сила постоянна и отрицательна, значит, движение равнозамедленное. В этом простейшем случае, когда сила в действительности от x не зависит, конечно, не было надобности применять такой сложный способ расчета.

В следующем примере (по поводу которого см. также гл. 16) рассмотрим потенциальную энергию, график которой имеет вид ступеньки (рис. 2, *а*). Такой зависимости $u(x)$ соответствует график силы, приведенный на рис. 2, *б* (чтобы убедиться в этом, читатель должен вспомнить, что $F = -\frac{du}{dx}$): сила весьма велика и отрицательна, т. е. направлена в сторону уменьшения x. Чем круче кривая $u(x)$ на рис. 2, *а*, т. е. чем на меньшем протяжении $\Delta = x_1 - x_0$ происходит подъем $u(x)$, тем больше по абсолютной величине сила. Там, где $u(x)$ постоянна (слева от точки x_0 и справа от точки x_1), сила равна нулю.

Пусть тело начало движение от точки x_0 (см. рис. 2, *а*) со скоростью v_0; полная энергия тела пусть равна E. При каких значениях E тело может попасть в точку x_1? Так как $u(x_0) = 0$, то $E = \frac{mv_0^2}{2}$. С другой стороны, $E = \frac{mv_1^2}{2} + u_1$, где v_1 — скорость тела в точке x_1, а u_1 — потенциальная энергия при $x = x_1$. Поэтому

$$\frac{mv_1^2}{2} = E - u_1. \qquad (10.1.7)$$

Из формулы (7) видно, что если $E < u_1$, то тело не может попасть в точку x_1, так как в этом случае получаем $v_1^2 < 0$, чего быть не может. Поэтому тело может попасть в точку $x = x_1$, только если $E \geqslant u_1$.

Определим для этого случая работу силы F при перемещении тела из точки x_0 в x_1:

$$A = \frac{mv_1^2}{2} - \frac{mv_0^2}{2} = \frac{mv_1^2}{2} - E.$$

Пользуясь (7), находим

$$A = E - u_1 - E = -u_1.$$

При дальнейшем движении тела вправо от точки x_1 сила F работы не совершает, так как $F = 0$ при $x > x_1$.

Упражнения

10.1.1. Потенциальная энергия дана формулой $u = \frac{kx^2}{2}$, где $k > 0$. Построив график, покажите, что соответствующее движение — колебательное.

10.1.2. Потенциальная энергия $u(x) = 0$ при $x \leqslant 0$; $u(x) = 2x$ при $0 \leqslant x \leqslant 1$; $u(x) = 2$ при $x \geqslant 1$. В начальный момент тело с массой 1 г выходит из начала координат и движется вправо со скоростью v_0 (см/с), где: а) $v_0 = 1$; б) $v_0 = 1{,}9$; в) $v_0 = 2{,}1$. Для каждого случая укажите, сможет ли тело неограниченно двигаться вправо. Если не может, то найдите точку остановки.

10.1.3. $u(x) = -x^3 + 4x^2$. В начальный момент времени тело массой 2 г выходит из точки x_0 и движется со скоростью v_0 (см/с),

где: а) $x_0=1$, $v_0=1$; б) $x_0=-2$, $v_0=1$; в) $x_0=-2$, $v_0=-1$. В каждом случае исследуйте характер движения (точки остановки, области, в которые тело не может попасть). В случаях, когда есть точки остановки, укажите, хотя бы грубо, их координаты.

10.1.4. Ответьте на те же вопросы, что и в упр. 3, для $u(x)=\frac{x^2}{1+x^2}$, $m=2$, где: а) $x_0=0$, $v_0=2$; б) $x_0=\frac{1}{2}$, $v_0=\frac{1}{2}$. Выразите время t в зависимости от x через интеграл.

§ 2. Случай силы, пропорциональной отклонению. Гармонические колебания

Рассмотрим тело, на которое действует сила

$$F=-kx.$$

Как мы знаем, такой силе соответствует потенциальная энергия

$$u=\frac{kx^2}{2}.$$

Начало координат является положением устойчивого равновесия. Кривая потенциальной энергии (парабола) имеет вид, показанный на рис. 1.1.

Движение тела под действием такой силы представляет собой *колебания* влево и вправо от положения равновесия. Можно представить себе шарик, который скатывается с одной ветки параболы, набирая скорость, по инерции забирается на вторую ветвь, скатывается с нее и т. д.[1] Согласно второму закону Ньютона уравнение этих колебаний имеет вид

$$m\frac{d^2x}{dt^2}=-kx. \qquad (10.2.1)$$

Мы не будем решать его общим (и довольно сложным) способом (см. § 1), а вместо этого попытаемся угадать вид решения, а затем обратимся к исследованию его свойств.

Итак, предположим, что

$$x=a\cos\omega t. \qquad (10.2.2)$$

Такой вид решения выбран потому, что интересующий нас физический процесс («колебания»), бесспорно, имеет периодический характер, а косинус является одной из простейших периодических функций. Подставим (2) в основное уравнение (1); так как

$$v=\frac{dx}{dt}=-a\omega\sin\omega t,$$

$$\frac{d^2x}{dt^2}=-a\omega^2\cos\omega t,$$

то получаем

$$-ma\omega^2\cos\omega t=-ka\cos\omega t. \qquad (10.2.3)$$

Если $m\omega^2=k$, то соотношение (3) будет выполняться при всех t. Поэтому функция (2) действительно удовлетворяет уравнению (1), если только $m\omega^2=k$, т. е.

$$\omega=\sqrt{\frac{k}{m}}. \qquad (10.2.4)$$

Таким образом, имеем

$$x=a\cos\left(t\sqrt{\frac{k}{m}}\right). \qquad (10.2.5)$$

Отметим, что квадратный корень в выражении для ω не приводит к двум решениям, так как $\cos\omega t=\cos(-\omega t)$.

Мы искали решение уравнения (1) в виде (2). Но ясно, что синус ничем не хуже косинуса[2], — и уравнение (1) имеет и другое решение

$$x(t)=b\sin\omega t. \qquad (10.2.2а)$$

В самом деле, $\frac{d^2(\sin\omega t)}{dt^2}=-\omega^2\sin\omega t$; подставляя значение $x(t)$ и его второй производной в (1) и сокращая на

---

[1] Разумеется, эта картина не является точным описанием интересующего нас процесса: ведь скорость скатывающейся с ветви параболы частицы направлена по касательной к параболе и может быть разбита на две компоненты: v_x (горизонтальная составляющая скорости) и v_z (вертикальная составляющая). В законе сохранения энергии, естественно, фигурируют обе величины v_x и v_z, в силу чего действующая на частицу сила (точнее, горизонтальная составляющая этой силы) не будет строго равна $-kx$ (подумайте сами, почему?). Однако в случае достаточно пологой параболы, когда составляющая v_z скорости невелика, «скатывание с параболы» можно представить себе как колебательное движение по закону (1).

[2] Обе функции $x_1=a\sin\omega t$ и $x_2=b\cos\omega t$ характеризуются пропорциональностью второй производной функции самой функции (с *отрицательным* коэффициентом пропорциональности: $x_1''=-\omega^2x_1$; $x_2''=-\omega^2x_2$); различаются же они *начальными условиями* (при $t=0$): в то время как $x_1(0)=0$, для функции x_2 равна нулю начальная скорость ее изменения ($x_2'(0)=0$), а не сама функция.

$\sin\omega t$, получим $\omega=\sqrt{\frac{k}{m}}$, т. е. то же самое значение, что и раньше.[3] Поэтому

$$x(t)=b\sin\left(t\sqrt{\frac{k}{m}}\right). \quad (10.2.5а)$$

Легко убедиться, что также и сумма

$$x=a\cos\omega t+b\sin\omega t \quad (10.2.6)$$

решений (2) и (2а) уравнения (1) является решением того же самого уравнения (1). (Читатель сам проверит это, находя вторую производную от суммы (6) и подставляя ее в (1); a и b здесь совершенно произвольны.) Таким образом, мы располагаем решением (6) уравнения (1) с двумя произвольными постоянными a и b.

Решение (6) уравнения (1) можно переписать и несколько по-другому. Рассмотрим «сдвинутое на угол φ» решение (2) уравнения (1):

$$x=C\cos(\omega t+\varphi), \quad (10.2.7)$$

где числовой множитель перед тригонометрической функцией мы теперь обозначаем буквой C. Так как и здесь[4] $x'=-\omega C\sin(\omega t+\varphi)$ и, значит, $x''=-\omega^2 C\cos(\omega t+\varphi)=-\omega^2 x$, то при выполнении (4) и функция (7) является решением (1). Но так как $\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta$, то (7) можно переписать так:

$$x=C\cos\varphi\cos\omega t-C\sin\varphi\sin\omega t,$$

т. е. (7) и (6) — это одно и то же решение, если только

$$a=C\cos\varphi, \quad b=-C\sin\varphi, \quad (10.2.8)$$

где C и φ выражаются через a и b[5]:

$$C=\sqrt{a^2+b^2}, \quad \operatorname{tg}\varphi=-\frac{b}{a},$$

$$\varphi=\operatorname{arc\,tg}\left(-\frac{b}{a}\right). \quad (10.2.8а)$$

Найдем *период* T колебаний, т. е. время, через которое тело возвращается в исходное положение (причем даже, как нетрудно видеть (см. формулы (10) и (11)), с исходной скоростью). Функция $\cos\alpha$ (или $\sin\alpha$, или $\cos(\alpha+\varphi)$) принимает первоначальное значение, когда угол α меняется на 2π. Значит, в выражении $a\cos\omega t$ величина ωt за один период T должна измениться на 2π. Поэтому T находим из условия

$$\omega(t+T)=\omega t+2\pi,$$

откуда

$$\omega T=2\pi, \quad T=\frac{2\pi}{\omega}=2\pi\sqrt{\frac{m}{k}}. \quad (10.2.9)$$

Величина $\nu=\frac{1}{T}$ дает число колебаний в единицу времени и называется *частотой* колебаний. Размерность ее $с^{-1}=1/с$ (обратная секунда). Единица частоты — одно колебание в секунду — имеет специальное название *герц* (обозначение: Гц) в честь великого немецкого физика Генриха Герца (1857—1894). Из формулы (9) видно, что $\nu=\frac{\omega}{2\pi}$. Однако во всех формулах удобнее иметь дело именно с ω, а не с ν, иначе повсюду появятся коэффициенты 2π и $4\pi^2$. Величина $\omega=\frac{2\pi}{T}$ называется *круговой частотой*[6], а φ в формуле (7) — *начальной фазой*.

---

[3] Здесь также $\omega=-\sqrt{\frac{k}{m}}$ не дает нового решения: $b\sin\left(-\sqrt{\frac{k}{m}}\,t\right)=-b\sin\left(\sqrt{\frac{k}{m}}\,t\right)$ и значение $\omega=-\sqrt{\frac{k}{m}}$ отвечает тому же решению (5а), но только с иным коэффициентом b.

[4] Ясно, что *сдвиг* аргумента тригонометрической функции не влияет на ее дифференциальные свойства: если $x=\cos(t+\alpha)=\cos\tau$, где $\tau=t+\alpha$ и $\alpha=\text{const}$, то $\frac{dx}{dt}=\frac{dx}{d\tau}\frac{d\tau}{dt}=\frac{dx}{d\tau}\cdot 1=\frac{dx}{d\tau}$. (Переход от (2) к (7) — это то же, что сдвиг по времени: $t\to t+t_0$, где $t_0=\frac{\varphi}{\omega}$.)

[5] Разумеется, для полного определения угла φ простой формулы $\varphi=\operatorname{arc\,tg}\left(-\frac{b}{a}\right)$ недостаточно — здесь надо еще учесть знаки функций $\cos\varphi$ и $\sin\varphi$, указываемые формулой (8).

[6] Чтобы понять происхождение этого названия, рассмотрим отрезок длины a, вращающийся против часовой стрелки. (Сходство между вращением и колебанием бросается в глаза: вращающаяся стрелка после каждого оборота возвращается в исходное положение точно так же, как колеблющееся тело по истечении одного периода возвращается в свое исходное положение.) При этом координата x конца вращающейся стрелки меняется по закону $x=a\cos\omega t$, если стрелка вращается с угловой скоростью ω. Если T есть время одного оборота (период вращения), то $\nu=\frac{1}{T}$ есть число оборотов в единицу времени, $\omega=2\pi\nu$ есть угловая скорость вращения, выраженная в радианах в секунду. Так как радиан есть величина безразмерная, то ω имеет размерность $1/с$. В связи с таким простым смыслом ω при движении по кругу величина ω

Коэффициент a решения (2) из уравнения (1) определить нельзя, потому что уравнение удовлетворяется при л ю б о м a (в обеих частях (3) a можно сократить). Аналогично совершенно произвольны константа b в решении (2а) и обе константы a и b в решении (6); также произвольны величины C и φ в формуле (7) (проверьте!).

Если движение тела описывается формулой (2), то скорость тела

$$v = \frac{dx}{dt} = -a\omega \sin \omega t. \qquad (10.2.10)$$

Из соотношения $\cos^2 \omega t + \sin^2 \omega t = 1$ следует, что при $\cos \omega t = \pm 1$ будет $\sin \omega t = 0$. Следовательно, в те моменты времени, когда отклонение тела в ту или иную сторону достигает наибольшего значения ($x=a$ или $x=-a$), скорость v равна нулю. Представим себе, что при $t < 0$ тело помещено в точку $x=a$ и удерживается в этой точке в покое при помощи посторонней силы (например, каким-нибудь крючком) до момента $t=0$, когда крючок отпускает тело. В этот момент тело было в покое, а затем под действием силы $F=-kx$ начались колебания. Зависимость координаты x тела от времени t дается формулой $x=a \cos \omega t$. Так как абсолютная величина $\cos \omega t$ не превосходит 1, то a есть наибольшее значение величины x, т. е. *наибольшее отклонение тела от положения равновесия.* Тот же смысл имеет величина b в формуле (2а) и C в формуле (7). Число a (соответственно b или C) называется *амплитудой* колебаний. Из (2) и (10) следует, что амплитуда колебаний равна начальному отклонению тела, если в момент начала колебаний тело покоилось [7].

Отметим попутно, что вообще если $A(t)=L \cos \omega t$ (или $A(t)=L \sin \omega t$), то L есть наибольшее значение величины $A(t)$; оно называется *амплитудой* величины $A(t)$. Укажем еще, что частота колебаний ω не зависит от их амплитуды a.

Пусть $x=x_1(t)$ есть решение уравнения (1), т. е. справедливо равенство $m\frac{d^2x_1}{dt^2} = -kx_1$. Рассмотрим функцию $x_2(t) = cx_1(t)$, где c — постоянная величина. Подставляя в уравнение (1) значения x_2 и $\frac{d^2x_2}{dt^2}$, получим $mc\frac{d^2x_1}{dt^2} = -kcx_1(t)$, или, сокращая на c,

$$m\frac{d^2x_1}{dt^2} = -kx_1.$$

Итак, если $x=x_1(t)$ удовлетворяет уравнению (1), то и $x_2(t)=cx_1(t)$ также удовлетворяет этому уравнению. Аналогично проверяется тот факт, что если $x_1(t)$ и $x_2(t)$ — два решения уравнения (1), то $x(t)=x_1(t)+x_2(t)$ тоже является решением. Именно так из решений (2) и (2а) образуется более общее решение (6). Соответствующая (6) скорость движения равна

$$v = \frac{dx}{dt} = -a\omega \sin \omega t + b\omega \cos \omega t. \qquad (10.2.11)$$

Допустим, что мы взяли решение $x=a \cos \omega t$ уравнения (1). Полагая $t=0$, получим $x_0=a$, где $x_0=x(0)$ — начальное положение тела и, значит, $x = x_0 \cos \omega t$. Но тогда $v = \frac{dx}{dt} = -x_0\omega \sin \omega t$, так что $v=0$ при $t=0$. Поэтому решение $x=a \cos \omega t$ годится лишь для задачи о колебании с *нулевой начальной скоростью.*

Попробуем взять решение $x=b \sin \omega t$. В этом случае $v = \frac{dx}{dt} = b\omega \cos \omega t$, и при $t=0$ получаем $v_0 = b\omega$, где $v_0 = v(0)$ — начальная скорость и

$$b = \frac{v_0}{\omega}, \qquad (10.2.12)$$

откуда $x = \frac{v_0}{\omega} \sin \omega t$. Однако при $t=0$ мы имеем здесь обязательно $x=0$, и при помощи этого решения мы также не можем решить общую задачу — неизбежно приходится считать, что колебание началось с точки равновесия $x=0$.

Соотношение (12) дает практически удобный способ измерения импульса силы и скорости, широко применяемый в механике под названием *баллистического маятника*: если тело подвешено в виде маятника или удерживается в положении равновесия пружинами и частота его колебаний известна, то начальную скорость после удара можно определить по амплитуде колебаний, вызванных ударом.

в задачах о колебаниях и получила название *круговой частоты.*

[7] Мы определили амплитуду как п о л о в и н у полного размаха колебаний. На пути от крайней левой точки $x=-a$ до крайней правой точки $x=a$ тело проходит расстояние $2a$, равное у д в о е н н о й амплитуде.

Покажем, что приближенно формулу (12) можно получить из элементарных соображений общего порядка. Размерность амплитуды — см, размерность скорости — см/с, а размерность периода колебаний — с. Поэтому соображения размерности подсказывают совпадение амплитуды с произведением начальной скорости на некоторую долю периода. Так как движение от момента удара до максимального отклонения продолжается четверть периода и $v < v_0$, потому что движение замедленно, то $b < \frac{v_0 T}{4}$. Если бы движение было *равнозамедленным*, то средняя скорость равнялась бы половине начальной; следовательно, можно рассчитывать, что $b \approx \frac{v_0 T}{8}$. В действительности же, как следует из формул (9) и (12),

$$b = \frac{v_0}{\omega} = \frac{v_0 T}{2\pi} = \frac{v_0 T}{6{,}28},$$

что достаточно близко к полученному «прикидкой» результату. Важно, что вследствие независимости периода колебаний от амплитуды последняя прямо пропорциональна начальной скорости.

Мы проверили, что две различные функции (2) и (2а) удовлетворяют уравнению $m\frac{d^2x}{dt^2} = -kx$. Пусть мы хотим решить задачу о движении тела с заданными начальным положением и начальной скоростью: $x = x_0$, $v = v_0$ при $t = 0$ (значения x_0 и v_0 — любые). Такую задачу назовем *общей*. До сих пор в отличие от общей задачи мы рассматривали лишь *частные* задачи. В одной из них при $t = 0$ было $x = x_0$, а $v = 0$; а в другой — при $t = 0$ было $v = v_0$, но $x = 0$.

При помощи (6) уже можно решить общую задачу о движении тела с произвольным положением и произвольной скоростью в начальный момент: $x = x_0$, $v = v_0$ при $t = 0$. Пользуясь начальными условиями, находим из (6) и (11): $a = x_0$, $b = \frac{v_0}{\omega}$. Поэтому

$$x = x_0 \cos \omega t + \frac{v_0}{\omega} \sin \omega t. \qquad (10.2.13)$$

Решение (6) называется *общим* решением уравнения (1), в то время как (2) и (2а) — это *частные* его решения.

Задаваемое формулой (6) (или (7)) движение называется *гармоническим колебанием*.

Если решение записано в форме (7), то ясно, что амплитуда колебаний равна C. Следовательно, если решение имеет вид (6), то амплитуда C равна $\sqrt{a^2 + b^2}$. Пусть $x = x_0$, $v = v_0$ при $t = 0$; тогда $a = x_0$, $b = \frac{v_0}{\omega}$, и поэтому амплитуда колебаний

$$C = \sqrt{x_0^2 + \frac{v_0^2}{\omega^2}}. \qquad (10.2.14)$$

Упражнение

10.2.1. Тело колеблется по закону $\frac{d^2x}{dt^2} = -x$. Найдите зависимость $x(t)$ и определите период колебаний для следующих случаев: а) $x = 0$, $v = 2$ см/с при $t = 0$; б) $x = 1$, $v = 0$ при $t = 0$; в) $x = 1$, $v = 2$ см/с при $t = 0$. В случае в) запишите решение как в виде (6), так и в виде (7).

§ 3. Маятник

Эталонной задачей, с которой в значительной степени связано учение о гармонических колебаниях, является задача о *маятнике*. Если представить себе (точечную) массу m, сосредоточенную на конце невесомой нити длины l, подвешенной за второй конец (рис. 1), то в случае отклонения нити (*математического маятника*) на угол φ на массу m действует сила тяжести, направленная вертикально вниз и равная mg, где g — ускорение свободного падения. Нормальная (т. е. направленная по нормали к траектории массы m) составляющая этой силы уравновешивается натяжением нити; тангенциальная же составляющая, направленная по касательной к траектории, равна $-mg \sin \varphi$; она-то и является причиной движения маятника. (Знак минус у тангенциальной составляющей силы объясняется тем, что ее направление совпадает с направлением убывания угла φ). Так как проходимый от положения равновесия 0 путь s равен $l\varphi$, то скорость массы m равна $l\frac{d\varphi}{dt}$, а ускорение равно $l\frac{d^2\varphi}{dt^2}$; поэтому второй закон Ньютона записывается так:

$$ml\frac{d^2\varphi}{dt^2} = -mg \sin \varphi, \quad \text{или}$$

$$\frac{d^2\varphi}{dt^2} = -\frac{g}{l} \sin \varphi. \qquad (10.3.1)$$

Точное решение уравнения (1) вовсе не просто (ср. с. 234). Однако, когда угол φ мал, величина $\sin\varphi$ может быть заменена первым членом φ разложения $\sin\varphi$ в ряд Тейлора (см. § 6.2); при этом уравнение (1) заменится следующим приближенным уравнением:

$$\frac{d^2\varphi}{dt^2}=-\frac{g}{l}\varphi. \qquad (10.3.1a)$$

Ясно, что (1а) — это уравнение (2.1) гармонических колебаний, где роль коэффициента $\frac{k}{m}$ правой части (если поделить обе части (2.1) на m) играет отношение $\frac{g}{l}$. Поэтому решение (1а) имеет вид (2.7), где $\omega=\sqrt{\frac{g}{l}}$. Период T колебаний маятника задается формулой

$$T=\frac{2\pi}{\omega}=2\pi\sqrt{\frac{l}{g}}. \qquad (10.3.2)$$

В частности, если потребовать, чтобы было $T=2$ с, т. е. чтобы качание в одну сторону маятник совершал за 1 с, то, положив $g=9{,}8$ м/с^2, получим $l=\frac{gt^2}{4\pi^2}\approx 0{,}993$ м, т. е. $l\approx 1$ м[1].

Вот иной подход к той же задаче о математическом маятнике; он будет нам удобен при рассмотрении более общей задачи о физическом маятнике. Пусть маятник при $x=0$ (положение равновесия) имеет потенциальную энергию u_0; кинетическая его энергия в этот момент равна нулю. Отклоним маятник на некоторый угол φ; в результате этого по горизонтали он отклонится на величину x (см. рис. 1). В этом положении потенциальная энергия маятника $u_1=u_0+mgz$, где $z=l-\sqrt{l^2-x^2}$; кинетическая энергия его равна $\frac{m}{2}\left(\frac{dx}{dt}\right)^2$ (ср. § 9.6). В процессе колебаний сумма кинетической и потенциальной энергий не изменяется, поэтому

$$u_0+mg\left(l-\sqrt{l^2-x^2}\right)+\frac{m}{2}\left(\frac{dx}{dt}\right)^2=u_0,$$

или

$$\left(\frac{dx}{dt}\right)^2=-2g\left(l-\sqrt{l^2-x^2}\right).$$

Теперь воспользуемся тем, что $x\ll l$ (колебания малые), т. е. $\frac{x}{l}\ll 1$. Благодаря этому $\sqrt{l^2-x^2}$ можно представить так:

$$\sqrt{l^2-x^2}=l\sqrt{1-\left(\frac{x}{l}\right)^2}\approx$$
$$\approx l\left[1-\frac{1}{2}\left(\frac{x}{l}\right)^2\right]=l-\frac{x^2}{2l}$$

(здесь мы оставляем всего два первых члена в разложении $\sqrt{1-\left(\frac{x}{l}\right)^2}=$

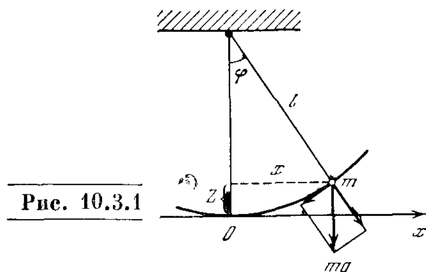


Рис. 10.3.1

$=\left[1-\left(\frac{x}{l}\right)^2\right]^{1/2}$ по степеням $\frac{x}{l}$ по формуле Маклорена; см. гл. 6), после чего уравнение колебаний маятника принимает вид

$$\left(\frac{dx}{dt}\right)^2=-g\frac{x^2}{l}.$$

Продифференцировав обе части последнего равенства по t, получим

$$2\frac{dx}{dt}\frac{d^2x}{dt^2}=-2g\frac{x}{l}\frac{dx}{dt},$$

откуда

$$\frac{d^2x}{dt^2}=-\frac{g}{l}x. \qquad (10.3.1б)$$

Это — иная форма уравнения малых колебаний (математического) маятника; сходство ее с (1а) объясняется тем, что при малых x и φ имеет место приближенное равенство $\sin\varphi=\frac{x}{l}\approx\varphi$, т. е. $x\approx l\varphi$.

Развитые соображения могут быть применены и к изучению колебательного движения твердого тела, закрепленного в определенной «точке

[1] При установлении во Франции в период Великой французской революции метрической системы мер в качестве основной единицы длины конкурировали длина одной сорокамиллионной части парижского меридиана (именно она и была названа метром и положена в основу новой системы мер) и длина подобного «секундного» маятника (зависящая, очевидно, от выбора точки земного шара, поскольку от этого зависит величина g); обе эти длины имеют один порядок величины, хорошо согласованный с характерными размерами человеческого тела.

подвеса». Для простоты мы будем говорить о стержне, подвешенном в некоторой точке A, удаленной от центра тяжести стержня на расстояние l (рис. 2). Определим период колебаний стержня (*физического маятника*).

Если маятник отклонен от положения равновесия на малый угол φ, то высота z центра тяжести над самым низким его положением, отвечающим равновесию стержня, равна $l - l\cos\varphi$ (ср. выше); поэтому потенциальная энергия u стержня может быть записана так:

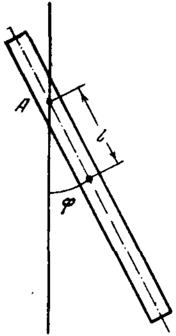


Рис. 10.3.2

$$u = mgz = mg(l - l\cos\varphi) = mgl(1-\cos\varphi). \qquad (10.3.3)$$

Разложим $\cos\varphi$ в ряд по степеням φ и, так как φ мало, ограничимся, как и ранее, первыми двумя членами ряда: $\cos\varphi \approx 1 - \frac{\varphi^2}{2}$. Поэтому приближенно имеем

$$u \approx mgl\frac{\varphi^2}{2}.$$

Ясно, что это соотношение указывает на увеличение потенциальной энергии при отклонении маятника на угол φ от положения равновесия, когда $\varphi=0$, $z=0$ и $u=0$.

Кинетическая энергия вращения стержня вокруг A равна

$$K = I\frac{\omega^2}{2} = I\cdot\frac{1}{2}\left(\frac{d\varphi}{dt}\right)^2, \qquad (10.3.4)$$

где I — момент инерции стержня (относительно точки его подвеса). Но согласно формуле (9.12.13) $I=ml^2+I_0$, поэтому

$$K = \frac{1}{2}(ml^2+I_0)\left(\frac{d\varphi}{dt}\right)^2,$$

где I_0 — момент инерции маятника относительно центра тяжести.

Предположим, что стержень совершает гармонические колебания, т. е. $\varphi = a\cos\omega t$. По закону сохранения энергии $u+K=\text{const}$ и $u_{max}=K_{max}$; значение K_{max} достигается в положении равновесия, когда $u=0$, а значение u_{max} — в положении, где $K=0$. Так как $\frac{d\varphi}{dt} = -a\omega\sin\omega t$, то

$$K_{max} = \frac{1}{2}(ml^2+I_0)a^2\omega^2, \quad u_{max} = mgl\frac{a^2}{2},$$

откуда

$$mgl\frac{a^2}{2} = \frac{1}{2}(ml^2+I_0)a^2\omega^2,$$

т. е.

$$\omega = \sqrt{\frac{mgl}{ml^2+I_0}}. \qquad (10.3.5)$$

Период колебаний равен

$$T = \frac{2\pi}{\omega}. \qquad (10.3.5a)$$

Из формул (5) и (5а) следует, что чем больше момент инерции I_0 стержня, тем меньше частота колебаний физического маятника, соответственно тем больше период T.

Если вся масса стержня сосредоточена в его центре тяжести, то $I_0=0$. В этом случае получаем выписанные выше хорошо известные формулы (2) для периода колебаний (и для частоты) математического маятника.

В случае, когда $I_0 \neq 0$, имеется определенное положение точки подвеса, при котором частота колебаний м а к с и м а л ь н а. Так как положение точки подвеса характеризуется величиной l, то для отыскания интересующего нас положения приравняем нулю производную $\frac{d\omega}{dl}$. После простых алгебраических преобразований мы придем к следующему уравнению для l:

$$mg(ml^2+I_0) - mgl\cdot 2ml = 0, \qquad (10.3.6)$$

откуда получаем

$$l_{max} = \sqrt{\frac{I_0}{m}}.$$

Для стержня длины L с р а в н о м е р н о распределенной массой $I_0 = \frac{mL^2}{12}$ (см. упр. 9.12.1), поэтому $l_{max} = \frac{L}{\sqrt{12}} \approx 0{,}3L$.

Заметим, что м и н и м а л ь н о е значение частоты $\omega=0$ достигается, очевидно, при $l=0$, т. е. при закреплении маятника в его центре тяжести, когда колебательное движение не имеет места [2]. Однако значение $l=0$ не является корнем уравнения (6), т. е. не обращает производную $\frac{d\omega}{dl}$ в нуль —

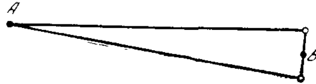
[2] Это значение $\omega=0$, конечно, условно: оно означает, что при $l\to 0$ и $\omega\to 0$ (чем ближе точка подвеса к центру тяжести стержня, тем больше период T колебаний).

здесь мы имеем место с «негладким» минимумом, достигаемым на границе области изменения переменной l (ср. § 7.2).

Упражнение

10.3.1. Маятник имеет вид тонкой треугольной полоски из листового материала (жесть, картон) (рис. 3). Определите период

Рис. 10.3.3



колебаний, если маятник подвешен: а) за острый конец A, б) за середину основания B. В обоих случаях определите, как надо сместить точку подвеса, чтобы получить минимальный период колебаний.

§ 4. Энергия колебаний. Затухающие колебания

Запишем общее решение уравнения (2.1) (общее гармоническое колебание) в виде

$$x = C\cos(\omega t + \alpha).$$

Потенциальная энергия тела в каждый момент равна

$$u(x(t)) = \frac{kx^2(t)}{2} = \frac{kC^2}{2}\cos^2(\omega t + \alpha),$$

а кинетическая энергия

$$K(t) = \frac{mv^2}{2} = \frac{m}{2}[-C\omega\sin(\omega t + \alpha)]^2 = \frac{mC^2\omega^2}{2}\sin^2(\omega t + \alpha).$$

Частота колебаний, как мы уже знаем, определяется формулой $\omega^2 = \frac{k}{m}$. Подставляя это значение ω^2 в выражение для кинетической энергии, получим

$$K(t) = \frac{kC^2}{2}\sin^2(\omega t + \alpha).$$

Таким образом, множитель перед тригонометрической функцией в выражениях потенциальной и кинетической энергий одинаков. Сами функции $\cos^2(\omega t+\alpha)$ и $\sin^2(\omega t+\alpha)$ также очень похожи одна на другую: каждая из них может быть получена из другой смещением по оси времени на величину $\Delta t = \frac{\pi}{2\omega}$ (рис. 1). Каждая из величин u и K колеблется от максимального значения до нуля, причем когда одна величина максимальна, другая равна нулю. Отметим, что функции $\cos^2(\omega t+\alpha)$ и $\sin^2(\omega t+\alpha)$ описывают колебания вокруг среднего значения, равного половине максимального. Это обстоятельство легко усмотреть как из рис. 1, так и из известных формул

$$\cos^2\beta = {}^1/_2(1+\cos 2\beta) = {}^1/_2 + {}^1/_2\cos 2\beta, \tag{10.4.1}$$

$$\sin^2\beta = {}^1/_2(1-\cos 2\beta) = {}^1/_2 - {}^1/_2\cos 2\beta.$$

Здесь ясно, что величина ${}^1/_2\cos 2\beta$ колеблется, принимая то положительные, то отрицательные значения, а ${}^1/_2$ представляет собой среднее значение величины $\cos^2\beta$. Сумма потенциальной и кинетической энергий, т. е. полная энергия системы,

$$E = K + u = \frac{kC^2}{2}[\cos^2(\omega t + \alpha) + \sin^2(\omega t + \alpha)] = \frac{kC^2}{2}, \tag{10.4.2}$$

как и следовало ожидать, постоянна.

Отметим, что если бы мы задались движением с частотой, не удовлетворяющей формуле $\omega = \sqrt{\frac{k}{m}}$, то при таком движении сумма потенциальной и кинетической энергий не была бы постоянна, максимальная кинетическая энергия не равнялась бы максимальной потенциальной энергии. Это неудивительно, так как колебания с частотой, отличающейся от $\omega = \sqrt{\frac{k}{m}}$, не удов-

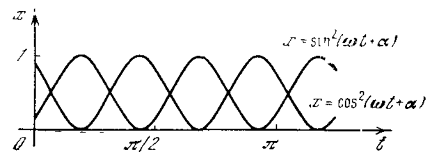


Рис. 10.4.1

летворяют основному уравнению движения; следовательно, для того чтобы такие колебания осуществлялись, необходимо, чтобы, кроме силы $F = -kx$ $\left(\text{ей соответствует потенциал } u(x) = \frac{kx^2}{2}\right)$, на тело действовали еще какие-то другие, внешние силы; за счет работы внешних сил полная энергия $\frac{mv^2}{2} + \frac{kx^2}{2}$ уже не будет сохранять свое значение.

Обратимся к вопросу о *затухании* колебаний. Пусть, кроме силы упругости пружины $F=-kx$, на тело действует единственная сила — сила трения. Предположим, что сила трения настолько невелика, что на протяжении одного периода колебаний работа этой силы мала по сравнению с энергией колебаний; тогда приближенно можно считать, что колебания происходят так

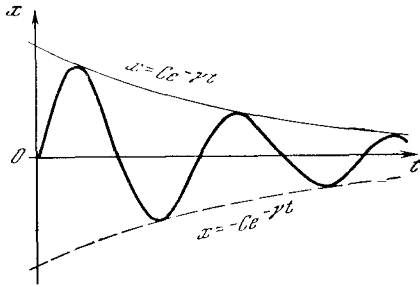


Рис. 10.4.2

же, как и в условиях отсутствия трения:

$$x(t)=C\cos(\omega t+\varphi)$$

(см. (2.7)). Энергия колебаний равна $\frac{kC^2}{2}$. Но наличие силы трения приведет к тому, что энергия колебаний с течением времени будет уменьшаться. Следовательно, действие трения скажется на том, что амплитуда C колебания (2.7) будет не постоянной, а медленно убывающей величиной. Закон убывания амплитуды C определится условием равенства уменьшения полной энергии E и работы силы трения F_1.

Относя обе величины к единице времени, получим

$$\frac{dE}{dt}=\frac{d\left(\frac{kC^2}{2}\right)}{dt}=kC\frac{dC}{dt}=F_1v=W_1, \tag{10.4.3}$$

где F_1 — сила трения, v — скорость тела, W_1 — мощность силы трения. Скорость v, так же как и сила F_1, в процессе колебаний периодически меняется. Для силы трения F_1 характерна зависимость ее от скорости v, причем F_1 по направлению всегда п р о т и в о п о л о ж н а v, так что произведение F_1v все время остается отрицательным. В интересующем нас случае м а л о г о трения, т. е. медленного затухания колебаний, можем считать, что изменение амплитуды $C(t)$ за время нескольких колебаний мало.

Под произведением F_1v надо понимать среднее значение этого произведения за период. Формула (3) справедлива только для промежутков времени, превышающих период колебаний.

Рассмотрим в качестве примера силу трения, *пропорциональную скорости* тела:

$$F_1=-hv,\quad F_1v=-hv^2. \tag{10.4.4}$$

Если здесь $v=\frac{dx}{dt}=-C\omega\sin(\omega t+\alpha)$, то

$$F_1v=-hC^2\omega^2\sin^2(\omega t+\alpha). \tag{10.4.5}$$

Заметим, что среднее значение $\sin^2(\omega t+\alpha)$ за период равно 1/2 (см. (1), а также § 7.7). Используя (3) и (4), получим окончательно

$$\frac{d}{dt}\left(\frac{kC^2}{2}\right)=kC\frac{dC}{dt}=-hC^2\omega^2\cdot\frac{1}{2},$$

откуда

$$\frac{dC}{dt}=-\frac{h\omega^2}{2k}C,$$

т. е. (ср. (6.6.10))

$$C=C_0e^{-\gamma t}, \tag{10.4.6}$$

где

$$\gamma=\frac{h\omega^2}{2k}. \tag{10.4.6а}$$

А поскольку $\omega^2=\frac{k}{m}$, то (6а) можно упростить:

$$\gamma=\frac{h}{2m}. \tag{10.4.6б}$$

В формуле (6) C_0 определяется из начальных условий. Умножая обе части (6) на $\cos(\omega t+\varphi)$ и используя (2.7), получим

$$x(t)=C_0e^{-\gamma t}\cos(\omega t+\varphi), \tag{10.4.7}$$

где $\omega=\sqrt{\frac{k}{m}}$ (см. характерный график затухающих колебаний на рис. 2). Это — приближенная формула, полученная в предположении, что сила трения невелика и пропорциональна скорости.

Во всех случаях, когда сила трения пропорциональна скорости, задача допускает точное решение. Мы считаем, что на тело действуют две силы: упру-

гая сила $-kx$ и сила трения $-h\frac{dx}{dt}$. В силу второго закона Ньютона

$$m\frac{d^2x}{dt^2}=-kx-h\frac{dx}{dt}. \qquad (10.4.8)$$

Решение $x(t)$ будем искать в таком же виде, как оно было получено для малой силы трения, т. е. в виде

$$x(t)=C_0e^{-\gamma t}\cos(\omega_1 t+\alpha). \qquad (10.4.9)$$

Решение (9) зависит от четырех параметров: C_0, γ, ω_1 и α. При этом значения γ и ω_1 определяются из уравнения (8), а C_0 и α остаются произвольными: законом (8) они не определяются, а находятся из начальных условий. В самом деле, из (9) получим:

$$\frac{dx}{dt}=-\gamma C_0e^{-\gamma t}\cos(\omega_1 t+\alpha)-$$
$$-C_0\omega_1e^{-\gamma t}\sin(\omega_1 t+\alpha),$$
$$\frac{dx^2}{dt^2}=\gamma^2C_0e^{-\gamma t}\cos(\omega_1 t+\alpha)+$$
$$+\gamma C_0\omega_1e^{-\gamma t}\sin(\omega_1 t+\alpha)+$$
$$+C_0\omega_1\gamma e^{-\gamma t}\sin(\omega_1 t+\alpha)-$$
$$-C_0\omega_1^2e^{-\gamma t}\cos(\omega_1 t+\alpha).$$

Подставляя эти выражения для x, $\frac{dx}{dt}$ и $\frac{d^2x}{dt^2}$ в (8) и сокращая на $C_0e^{-\gamma t}$, найдем

$$m\gamma^2\cos(\omega_1 t+\alpha)+m\gamma\omega_1\sin(\omega_1 t+\alpha)+$$
$$+m\gamma\omega_1\sin(\omega_1 t+\alpha)-m\omega_1^2\cos(\omega_1 t+\alpha)=$$
$$=-k\cos(\omega_1 t+\alpha)+h\gamma\cos(\omega_1 t+\alpha)+$$
$$+h\omega_1\sin(\omega_1 t+\alpha),$$

или

$$(m\gamma^2-m\omega_1^2)\cos(\omega_1 t+\alpha)+$$
$$+2m\gamma\omega_1\sin(\omega_1 t+\alpha)=$$
$$=(-k+h\gamma)\cos(\omega_1 t+\alpha)+$$
$$+h\omega_1\sin(\omega_1 t+\alpha). \qquad (10.4.10)$$

Равенство (10) выполняется при любом t, если

$$m\gamma^2-m\omega_1^2=-k+h\gamma,$$
$$2m\gamma\omega_1=h\omega_1. \qquad (10.4.11)$$

Из второго уравнения (11), сократив на ω_1, получим

$$\gamma=\frac{h}{2m}.$$

Теперь из первого уравнения (11) можно найти ω_1:

$$\omega_1^2=\frac{k}{m}-\frac{h^2}{4m^2}. \qquad (10.4.12)$$

Следовательно,

$$x(t)=C_0e^{-\frac{h}{2m}t}\cos\left(\sqrt{\frac{k}{m}-\frac{h^2}{4m^2}}\,t+\alpha\right). \qquad (10.4.13)$$

В случае, когда трение мало, т. е. когда h мало по сравнению с k, в (13) можно под корнем пренебречь членом $\frac{h^2}{4m^2}$ по сравнению с $\frac{k}{m}$. При этом (13) переходит в (7). Следовательно, в приближенном рассмотрении мы верно получили закон уменьшения амплитуды колебаний, но не заметили, что сила трения вызывает также и малое изменение их частоты. Укажем еще, что из двух законов (6), (6а) и (6), (6б) уменьшения амплитуды верным при всех силах трения (не слишком больших, т. е. таких, что $h^2<4km$ и корень в (13) существует) оказывается именно второй; первый из них совпадает со вторым при малых k, но в общем случае непригоден, поскольку в нем фигурирует (вообще говоря, не постоянная!) величина ω.

Если трение очень велико, т. е. $h^2>4km$, то подкоренное выражение в (13) становится отрицательным и формула (13) теряет смысл. Это означает, что при сильном трении движение уже не имеет колебательного характера. В этом случае решение уравнения (8) надо искать в виде $x=Ce^{-\gamma t}$. Подставив такое значение x в (8), мы получим квадратное уравнение для γ, имеющее два решения: $\gamma=\gamma_1$ и $\gamma=\gamma_2$. Сумма $C_1e^{-\gamma_1 t}+C_2e^{-\gamma_2 t}$ двух решений, соответствующих этим γ, даст нам общее решение (8) и позволит решить задачу с любыми начальными условиями. Подробно этот случай большого h будет рассмотрен в § 13.10 в связи с электрическими колебаниями.

Упражнения

10.4.1. Найдите закон затухания колебаний при силе трения, пропорциональной квадрату скорости (такое трение характерно для быстрого движения тела в жидкости с малой вязкостью). Покажите, что по истечении большого промежутка времени амплитуда

$C(t)=\frac{1}{bt}$, где b — постоянное число, не зависящее от C_0 — значения амплитуды в начальный момент времени.

10.4.2. Найдите закон затухания для силы трения, не зависящей от скорости (такая сила трения характерна для трения сухих твердых поверхностей одна о другую). Определите время, по истечении которого колебания прекратятся.

§ 5. Вынужденные колебания и резонанс

Рассмотрим тело, на которое действует упругая сила $F=-kx$. Мы установили, что под действием этой силы тело колеблется с определенной частотой $\omega=\sqrt{\frac{k}{m}}$, так называемой частотой свободных колебаний, или *собственной частотой*. В дальнейшем мы будем обозначать собственную частоту $\sqrt{\frac{k}{m}}$ колебаний через ω_0.

Пусть на тело, кроме упругой силы, действует еще и периодическая в н е ш н я я сила с частотой ω, вообще говоря не равной ω_0. Тогда оказывается, что амплитуда колебаний, вызванных внешней силой, весьма сильно зависит от того, насколько частота внешней силы близка к собственной частоте (т. е. к частоте свободных колебаний, происходящих в отсутствие внешних сил). Явление возрастания амплитуды при приближении ω к ω_0 носит название *резонанса* и имеет очень большое практическое значение; оно относится к любым системам, в которых возможны колебания. В механических системах (станках, моторах) такие колебания могут приводить к опасным деформациям и разрушению механизмов. Впрочем, явление резонанса может быть и полезным: иногда его сознательно используют для того, чтобы малой силой вызвать сильные колебания рабочего инструмента (колебания с большой амплитудой).

В электрических системах резонанс дает возможность при действии нескольких периодических сил разной частоты (например, нескольких передающих радиостанций) добиться того, чтобы колебания в нашей системе практически зависели только от о д н о й из этих сил, а именно от той, частота которой близка к собственной частоте системы. Благодаря этому можно настраивать радиоприемник на определенную станцию.

Составим уравнение колебаний:

$$m\frac{d^2x}{dt^2}=-kx-h\frac{dx}{dt}+f\cos\omega t, \quad (10.5.1)$$

где члены $-kx$ и $-h\frac{dx}{dt}$ отвечают упругой силе и трению (ср. (4.8)), а член $f\cos\omega t$ отражает действие внешней силы. Разделим обе части уравнения (1) на m и обозначим $\frac{k}{m}=\omega_0^2$ в соответствии с тем, что ω_0 есть частота собственных колебаний тела (в отсутствие трения); отношение $\frac{h}{m}$ обозначим через 2γ (ср. (4.6б)). Тогда (1) примет вид

$$\frac{d^2x}{dt^2}=-\omega_0^2x-2\gamma\frac{dx}{dt}+\frac{f}{m}\cos\omega t. \quad (10.5.1а)$$

Естественно ожидать, что под действием силы с частотой ω тело будет совершать колебания с той же частотой. Поэтому решение уравнения (1) или (1а) мы будем искать в следующем виде:

$$x=a\cos\omega t+b\sin\omega t. \quad (10.5.2)$$

Подставляя это выражение для x и выражения для производных x в уравнение (1а), получим

$$-a\omega^2\cos\omega t-b\omega^2\sin\omega t=$$
$$=-a\omega_0^2\cos\omega t-b\omega_0^2\sin\omega t+2\gamma a\omega\sin\omega t-$$
$$-2\gamma b\omega\cos\omega t+\frac{f}{m}\cos\omega t.$$

Для того чтобы последнее равенство выполнялось при в с е х t, надо, чтобы были в отдельности равны члены с $\cos\omega t$ и с $\sin\omega t$ в его левой и правой частях. Таким образом, мы приходим к следующей системе уравнений:

$$\left.\begin{aligned}-a\omega^2&=-a\omega_0^2-2\gamma b\omega+\frac{f}{m},\\ -b\omega^2&=-b\omega_0^2+2\gamma a\omega\end{aligned}\right\} \quad (10.5.3)$$

(ср. (4.11)). Из второго уравнения (3) следует, что

$$b=\frac{2\gamma\omega}{\omega_0^2-\omega^2}a.$$

Подставляя это выражение для b в первое уравнение (3), находим

$$a=\frac{f}{m}\frac{\omega_0^2-\omega^2}{(\omega_0^2-\omega^2)^2+(2\gamma\omega)^2}, \quad (10.5.4а)$$

а, значит,

$$b = \frac{f}{m}\frac{2\gamma\omega}{(\omega_0^2-\omega^2)^2+(2\gamma\omega)^2}. \qquad (10.5.4б)$$

Переходя теперь к записи колебаний $x = C\cos(\omega t + \alpha)$ и вспоминая, что $C = \sqrt{a^2+b^2}$ (см. (2.8а)), мы найдем амплитуду C колебаний, вызванных внешней силой:

$$C = \frac{f}{m}\frac{1}{\sqrt{(\omega_0^2-\omega^2)^2+(2\gamma\omega)^2}}. \qquad (10.5.5)$$

Мы видим, что C тем больше, чем ближе ω к ω_0. Кривая зависимости C от ω при данном ω_0 изображена на рис. 1 (для двух значений γ), где принято $\frac{f}{m} = 1$, $\omega_0 = 1$. Чем меньше трение h, т. е. чем меньше $\gamma = \frac{h}{2m}$, тем резче выражен подъем амплитуды колебаний при равенстве частоты внешней силы и собственной частоты.

Нетрудно убедиться, что сумма решения (4.9) уравнения (4.8) и общего решения (2) уравнения (1)

$$x = a\cos\omega t + b\sin\omega t + \\ + C_0 e^{-\gamma t}\cos(\omega_1 t + \varphi), \qquad (10.5.6)$$

где a и b даются формулами (4а) и (4б), а ω_1 и γ — формулами (4.12) и (4.6б), также является решением уравнения (1). Формула (6) позволяет найти решение, удовлетворяющее любым начальным условиям, для чего надо только подобрать соответствующие C_0 и φ. Действительно, пусть при $t=0$ имеем $x=x_0$, $v=v_0$. Подставляя эти значения в (6), находим

$$x_0 = a + C_0\cos\varphi, \\ v_0 = b\omega - C_0(\gamma\cos\varphi + \omega_1\sin\varphi). \qquad (10.5.7)$$

Из системы уравнений (7) можно определить C_0 и φ (см. упр. 1).

Таким образом, (6) есть *общее решение* задачи о колебаниях тела под действием упругой силы и периодической внешней силы. Это общее решение подтверждает сделанное выше предположение о том, что при длительном воздействии внешней силы с частотой ω тело будет колебаться с той же частотой ω. В самом деле, каковы бы ни были начальные условия, они влияют только на значения C_0 и φ, т. е. только на последнее слагаемое решения (6). Однако с течением времени это слагаемое, отвечающее частоте ω_0, затухает, становится сколь угодно малым по амплитуде (за счет стремящегося к нулю множителя $e^{-\gamma t}$) — и при больших t

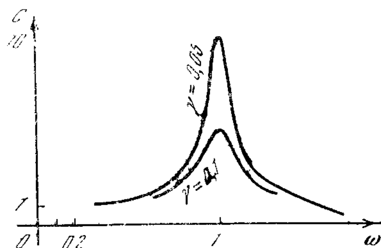


Рис. 10.5.1

им вполне можно пренебречь. Оставшиеся слагаемые описывают колебания с частотой ω, не затухающие с течением времени, поскольку они поддерживаются действием внешней силы, о которой предположено, что она имеет постоянную амплитуду.

Упражнения

10.5.1. Определите из системы (7) C_0 и φ.

10.5.2. Благодаря наличию трения максимальная амплитуда C получается при ω^2_{max}, несколько отклоняющемся от ω_0^2. Определите, насколько при равных γ отношение $\frac{\omega^2_{max}}{\omega_0^2}$ отклоняется от 1. [Указание. Исследуйте на минимум подкоренное выражение в (5), обозначив $\omega^2 = z$.]

§ 6. О точных и приближенных решениях физических задач

В предыдущем параграфе нам посчастливилось сравнительно просто найти точное решение задачи о колебаниях тела под действием периодической внешней силы при наличии упругой силы, стремящейся вернуть тело в состояние равновесия $(-kx)$, и силы трения $\left(-h\frac{dx}{dt}\right)$. Располагая этим точным решением, можно легко найти ряд его важных предельных случаев.

1) Пусть частота ω внешней силы *весьма мала* по сравнению с собственной частотой ω_0 $\left(где\ \omega_0^2 = \frac{k}{m}\right)$. Пренебрегая в (5.5) ω по сравнению с ω_0, мы получим

$$C = \frac{f}{m\omega_0^2} = \frac{f}{k}. \qquad (10.6.1)$$

2) Если, напротив, частота внешней силы ω *весьма велика* по сравнению с ω_0,

то можно отбросить в (5.5) ω_0 и считать, что

$$C=\frac{f}{m}\frac{1}{\sqrt{\omega^4+4\gamma^2\omega^2}}. \qquad (10.6.2)$$

Но $\gamma^2\omega^2 \ll \omega^4$ (трение не очень велико, так как иначе нет колебаний); пренебрегая в (2) членом $4\gamma^2\omega^2$, получаем

$$C=\frac{f}{m\omega^2}. \qquad (10.6.2а)$$

3) Если *сила трения h мала*, то в (5.5) можно пренебречь членом, содержащим γ; тогда мы получим

$$C=\frac{f}{m}\frac{1}{|\omega_0^2-\omega^2|}. \qquad (10.6.3)$$

(Появление абсолютной величины в (3) связано с тем, что в (5.5) подразумевается положительное значение корня: ведь ясно, что C и f, т. е. амплитуда колебания и амплитуда внешней силы, могут быть только положительны.)

4) Пусть имеет место явление точного *резонанса*: частота ω внешней силы в точности равна собственной частоте ω_0. Тогда (5.5) обращается в

$$C=\frac{f}{m}\frac{1}{2\gamma\omega}=\frac{f}{h\omega}=\frac{f}{h\omega_0}. \qquad (10.6.4)$$

Совокупность этих четырех предельных случаев, в сущности, составляет больше 90% содержания всех полученных результатов. Получив общий результат, всегда необходимо стараться его упростить, рассматривая различные предельные случаи общей ситуации. Простые формулы, относящиеся к предельным случаям, легче запоминаются и чаще используются на практике, в то время как к общим формулам приходится прибегать лишь изредка. Хотя предельные случаи и не охватывают полностью более сложную точную формулу, но практически они обычно дают о ней почти полную информацию.

Естественно возникает вопрос о том, нельзя ли было эти предельные формулы получить прямо, упрощая само уравнение, а не его решение. Ведь ясно, что решать точно сложное уравнение и лишь затем упрощать решение столь же неразумно, как с затратой большого труда аккуратно упаковать вместе несколько предметов, а затем разорвать упаковку, поскольку все равно этими предметами приходится пользоваться в отдельности. Поэтому очень важно научиться извлекать из общего уравнения задачи его частные (и предельные) случаи, не решая самого уравнения.

Прямое получение предельных (приближенных) выражений особенно важно еще и потому, что точное решение очень чувствительно к малейшим изменениям постановки задачи. Достаточно немного усложнить задачу — и точное решение уже не удается найти. Приближенное решение более грубо, но и более устойчиво относительно изменений условий задачи.

Особенно важны для учащегося те случаи, когда представляется возможным получить и сравнить между собой оба решения — точное и приближенное. Именно в таких случаях можно приобрести опыт правильного выбора приближений и уверенность в получаемых результатах: ведь правильность получаемых из общих формул предельных случаев может служить глобальной проверкой всего процесса решения общей задачи — и составления описывающего интересующий нас процесс уравнения и решения этого уравнения.

Вернемся к предельному случаю 1), когда частота ω внешней силы очень мала. Очевидно, что в этом случае мы имеем дело с весьма медленным движением. Поэтому в исходном уравнении (5.1)

$$m\frac{d^2x}{dt^2}=-kx-h\frac{dx}{dt}+f\cos\omega t$$

можно опустить члены $m\frac{d^2x}{dt^2}$ и $h\frac{dx}{dt}$, зависящие от скорости и ускорения движения: ведь если $x\approx C\cos(\omega t+\varphi)$, то $\frac{dx}{dt}$ пропорционально ω, а $\frac{d^2x}{dt^2}$ пропорционально ω^2 — и при малом ω эти величины будут малы. В таком случае мы получим

$$0\approx -kx+f\cos\omega t,$$

откуда

$$x\approx\frac{f\cos\omega t}{k}=C\cos\omega t, \quad \text{где } C=\frac{f}{k}. \qquad (10.6.5)$$

Таким образом, при малой частоте внешней силы в каждый момент времени приложенная внешняя сила $f\cos\omega t$ уравновешивается упругой силой kx. Ясно, что этот результат является весьма общим: он относится к любому движению с малой частотой.

Такой предельный случай называется *статическим*. В частности, сила упругости $G(x)$ может быть любой функцией координаты, а не только линейной функцией ($G=-kx$); внешняя сила $F(t)$ может быть любой функцией времени. Уравнение колебаний принимает тогда вид

$$m\frac{d^2x}{dt^2}=G(x)-h\frac{dx}{dt}+F(t). \qquad (10.6.6)$$

Точное решение этого уравнения можно получить далеко не всегда, но приближенный подход сохраняет силу во всех случаях. Действительно, пренебрегая в случае медленного движения членами (6), пропорциональными скорости $\frac{dx}{dt}$ и ускорению $\frac{d^2x}{dt^2}$, получаем

$$G(x)+F(t)\approx 0, \text{ т. е. } G(x)\approx -F(t).$$

Отсюда находим приближенную зависимость $x\approx x(t)$ между x и t. Подставляя это $x(t)$ в точное уравнение (6), можно найти, какого порядка ошибку мы допустили, пренебрегая членами $m\frac{d^2x}{dt^2}$ и $h\frac{dx}{dt}$.

Далее обратимся к предельному случаю 2), когда частота ω очень велика. При большой частоте время действия внешней силы, а следовательно, и импульс силы за каждый полупериод (пока сила действует в одном направлении) малы, потому что мала длительность полупериода. Значит, при данной амплитуде f внешней силы чем больше ω, тем меньше скорость, которую может набрать тело, и тем меньше перемещение тела. Пренебрегая в уравнении (5.1) членами kx и $h\frac{dx}{dt}$, получим уравнение движения свободного тела, на которое не действуют никакие силы, кроме внешней:

$$m\frac{d^2x}{dt^2}\approx f\cos\omega t. \qquad (10.6.7)$$

Будем искать решение уравнения (7) в виде $x=B\cos\omega t$. Тогда $\frac{d^2x}{dt^2}=-B\omega^2\cos\omega t$, и потому (7) принимает вид

$$-Bm\omega^2\cos\omega t\approx f\cos\omega t,$$

откуда

$$B\approx-\frac{f}{m\omega^2}$$

и, значит,

$$x\approx-\frac{f}{m\omega^2}\cos\omega t. \qquad (10.6.8)$$

Решение (8) можно записать в стандартном виде $x\approx C\cos(\omega t+\varphi)$, где C положительно:

$$x\approx\frac{f}{m\omega^2}\cos(\omega t+\pi). \qquad (10.6.8а)$$

При этом упругая сила есть

$$-kx\approx\frac{kf}{m\omega^2}\cos\omega t=\frac{\omega_0^2}{\omega^2}f\cos\omega t, \qquad (10.6.9)$$

а сила трения

$$-h\frac{dx}{dt}\approx-\frac{hf}{m\omega}\sin\omega t. \qquad (10.6.9а)$$

При сравнении сил, периодически зависящих от времени, нужно сравнивать не мгновенные их значения, а амплитуды. Отношение внешней силы $f\cos\omega t$ к упругой силе (9) — отношение амплитуд этих сил — здесь равно

$$f:\frac{\omega_0^2}{\omega^2}f=\frac{\omega^2}{\omega_0^2}.$$

Это отношение тем больше, чем больше ω. Аналогично и отношение внешней силы к силе трения (9а)

$$f:\frac{hf}{m\omega}=\frac{m}{h}\omega$$

неограниченно растет с ростом ω. Поэтому при больших ω внешняя сила значительно превосходит как упругую силу, так и силу трения. Этим подтверждается возможность приближенного рассмотрения движения под действием одной только внешней силы[1].

Предельный случай 3), когда трение пренебрежимо мало, является совсем простым. Здесь (5.1) принимает вид

$$m\frac{d^2x}{dt^2}\approx-kx+f\cos\omega t=$$
$$=-m\omega_0^2x+f\cos\omega t, \qquad (10.6.10)$$

ибо $\omega_0^2=\frac{k}{m}$ и, значит, $k=m\omega_0^2$. Решение уравнения (10) ищем в виде $x=C\cos(\omega t+\varphi)$. Подставляя в (10) выражения для x и $\frac{d^2x}{dt^2}$, получаем

$$\varphi=0,\ \pi\quad m\,|\,\omega_0^2-\omega^2\,|\,C\cos\omega t\approx f\cos\omega t,$$

[1] Существенно, что рассмотренная выше сила трения тем ближе к нулю, чем ближе скорость к нулю. При с у х о м т р е н и и (сила трения не зависит от скорости) внешняя сила, меньшая, чем сила трения, не вызовет колебаний н и п р и к а к о й частоте.

откуда

$$C \approx \frac{f}{m\,|\omega_0^2 - \omega^2|}. \qquad (10.6.11)$$

Сравнение (11) с точной формулой (5.5) позволяет оценить условия применимости этого приближения.

Ясно, что при $\omega = \omega_0$ приближение (11) неприменимо и предельный случай 4), когда частота внешней силы в точности равна собственной частоте колебаний (резонанс), следует рассматривать особо. (Однако трением мы теперь не пренебрегаем.) Будем искать решение уравнения (5.1) в стандартном виде $x=C\cos(\omega_0 t+\varphi)$. Тогда

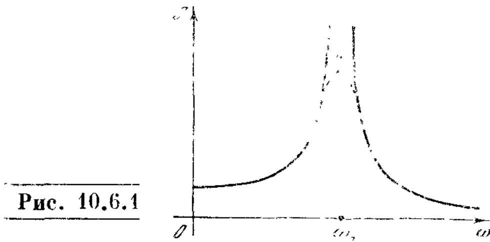


Рис. 10.6.1

$$m\frac{d^2x}{dt^2} = -mC\omega_0^2\cos(\omega_0 t + \varphi),$$

или, поскольку $\omega_0^2 = \frac{k}{m}$,

$$m\frac{d^2x}{dt^2} = -kC\cos(\omega_0 t + \varphi) = -kx.$$

Подставляя в (5.1) выражения для x и его производных, получим

$$hC\omega_0\sin(\omega_0 t + \varphi) + f\cos\omega_0 t = 0.$$

Последнее равенство справедливо при любом t, если $C = \left|-\frac{f}{h\omega_0}\right|$, или, точнее, если

$$C = \frac{f}{h\omega_0} \text{ и } \varphi = -\frac{\pi}{2}.$$

Следовательно, решение нашего уравнения таково:

$$x = \frac{f}{h\omega_0}\cos\left(\omega_0 t - \frac{\pi}{2}\right). \qquad (10.6.12)$$

Амплитуда колебаний при резонансе $C = \frac{f}{h\omega_0}$. Ясно, что пренебрегать трением h при резонансе никак нельзя!

Изобразим на чертеже зависимость C от ω, даваемую приближенными формулами (3) и (4) (рис. 1). Формула (3) дает две ветви кривой $C=C(\omega)$, уходящие в бесконечность при $\omega=\omega_0$; формула (4) при $\omega=\omega_0$ дает конечное значение $C=C_0\left(-\frac{f}{h\omega_0}\right)$. Построив кривую (3) и изобразив на том же чертеже точку $A=A(\omega_0, C_0)$, отвечающую формуле (4), уже нетрудно от руки провести гладкую кривую (пунктир на рис. 1), которая вдали от резонанса совпадает с кривыми (3) и имеет максимум C_0 в точке $\omega=\omega_0$.

Амплитуду C в случае резонанса можно определить при помощи *энергетических* соображений, ценность которых заключается в том, что они позволяют приближенно решить также и некоторые задачи, не имеющие точного решения. М о щ н о с т ь, развиваемая внешней силой $f\cos\omega t$, при движении, заданном формулой $x=C\cos(\omega t+\varphi)$, равна

$$W_{\text{вн}} = f\cos\omega t\frac{dx}{dt} =$$
$$= -fC\omega\cos\omega t\sin(\omega t + \varphi).$$

Определим с р е д н ю ю м о щ н о с т ь внешней силы за большой (точнее говоря, за бесконечный) промежуток времени:

$$\overline{W}_{\text{вн}} = -fC\omega\,\overline{\cos\omega t\sin(\omega t + \varphi)},$$

где черта сверху является знаком осреднения (см. § 7.7). Заметим, что

$$\cos\omega t\sin(\omega t+\varphi) = {}^1/_2\sin 2\omega t\cos\varphi + $$
$$+\cos^2\omega t\sin\varphi;$$

поэтому

$$\overline{\cos\omega t\sin(\omega t+\varphi)} = {}^1/_2\,\overline{\sin 2\omega t}\cos\varphi +$$
$$+\overline{\cos^2\omega t}\sin\varphi = {}^1/_2\sin\varphi,$$

так как $\overline{\sin 2\omega t} = 0$ и $\overline{\cos^2\omega t} = {}^1/_2$. Следовательно,

$$\overline{W}_{\text{вн}} = -\frac{fC\omega}{2}\sin\varphi,$$

что можно также записать как

$$\overline{W}_{\text{вн}} = \frac{fC\omega}{2}\cos\left(\varphi + \frac{\pi}{2}\right). \qquad (10.6.13)$$

Теперь определим среднюю мощность силы трения. Так как $F_{\text{тр}} = -hv$, то

$$\overline{W}_{\text{тр}} = -h\overline{v^2}. \qquad (10.6.14)$$

Но

$$\overline{v^2} = \overline{\left(\frac{dx}{dt}\right)^2} = C^2\omega^2\,\overline{\sin^2(\omega t+\varphi)} = \frac{C^2\omega^2}{2}$$

(ср. (4.1)). Поэтому (14) дает

$$\bar{W}_{\text{тр}} = -h\frac{C^2\omega^2}{2}.$$

Так как работа внешней силы идет на преодоление трения, то средние мощности внешней силы и силы трения должны быть равны:

$$\bar{W}_{\text{тр}} = \bar{W}_{\text{вн}}, \qquad (10.6.15)$$

т. е.

$$\frac{fC\omega}{2}\cos\left(\varphi+\frac{\pi}{2}\right) = -h\frac{C^2\omega^2}{2},$$

или

$$f\left|\cos\left(\varphi+\frac{\pi}{2}\right)\right| = hC\omega,$$

откуда

$$C = \frac{f}{h\omega}\left|\cos\left(\varphi+\frac{\pi}{2}\right)\right|. \qquad (10.6.16)$$

Максимальная возможная амплитуда (явление резонанса) получается, как видно из (16), при $\cos\left(\varphi+\frac{\pi}{2}\right)=1$, т. е. при $\varphi = -\frac{\pi}{2}$, чем определяется начальная фаза φ колебания. При этом $\omega=\omega_0$ и $C=\frac{f}{h\omega_0}$. Следовательно, в случае резонанса решение уравнения движения имеет вид

$$x = \frac{f}{h\omega_0}\cos\left(\omega_0 t - \frac{\pi}{2}\right),$$

т. е. мы опять получили формулу (12).

Вернемся к формуле (13). Из нее видно, что при резонансе $\bar{W}_{\text{вн}}$ имеет наибольшее значение, так как при резонансе $\cos\left(\varphi+\frac{\pi}{2}\right)=1$. Поэтому в случае резонанса внешняя сила развивает наибольшую среднюю мощность и, следовательно, производит наибольшую работу.

Приведенные энергетические соображения позволяют определить амплитуду при резонансе и в случае более сложной зависимости силы трения от скорости. Пусть сила трения пропорциональна фиксированной степени ви абсолютной величины скорости:

$$F_{\text{тр}} = -hv\,|v|^{n-1}, \qquad (10.6.17)$$

где $h>0$, и потому первый множитель гарантирует разные знаки v и $F_{\text{тр}}$. При $v>0$ (17) дает $F_{\text{тр}}=-hv^n$, а при $v<0$ получим $F_{\text{тр}}=h|v|^n$, т. е. во всех случаях v и $F_{\text{тр}}$ направлены в разные стороны. Поскольку мы по-прежнему считаем, что движение в основном порождается периодической внешней силой $F=f\cos\omega t$, т. е. имеет характер колебания $x=C\cos(\omega t+\varphi)$ частоты ω, то выражение (13) для средней мощности внешней силы справедливо и для этого случая.

Определим $\bar{W}_{\text{тр}}$. Мгновенное значение $W_{\text{тр}}=F_{\text{тр}}v=-hv^2|v|^{n-1}$; так как $v^2=|v|^2$, то $W_{\text{тр}}=-h|v|^{n+1}$. Подставляя сюда значение v, находим

$$W_{\text{тр}} = -hC^{n+1}\omega_0^{n+1}\,|\sin(\omega_0 t+\alpha)|^{n+1}. \qquad (10.6.18)$$

Из (18) следует[2]:

$$\bar{W}_{\text{тр}} = -hC^{n+1}\omega_0^{n+1}A, \quad \text{где}$$

$$A = \overline{|\sin(\omega_0 t+\alpha)|^{n+1}}.$$

Условие (15) дает

$$hC^{n+1}\omega_0^{n+1}A = \frac{1}{2}fC\omega_0\left|\cos\left(\alpha+\frac{\pi}{2}\right)\right|,$$

откуда

$$C = \sqrt[n]{\frac{f}{2hA\omega_0^n}\left|\cos\left(\alpha+\frac{\pi}{2}\right)\right|}.$$

Максимальная амплитуда, достигаемая при резонансе, равна

$$C = \frac{1}{\omega_0}\sqrt[n]{\frac{f}{2hA}}. \qquad (10.6.19)$$

Частным случаем формулы (19), отвечающим значению $n=1$ (сила трения пропорциональна скорости), является найденная раньше формула $C=\frac{f}{\omega_0 h}$.

Отметим, что при $n\neq 1$ уравнение движения становится нелинейным, и мы не можем выписать аналогичное найденному в § 5 общее решение этого уравнения: здесь точное решение элементарными методами найдено быть не может.

§ 7. Сложение колебаний. Биения

Довольно обычной является ситуация, когда одно и то же тело участвует в **нескольких** колебаниях. Так, видимый свет представляет собой суммарный эффект множеств колебательных процессов, связанных с перестройками отдельных атомов излучающего тела (ср. ниже, гл. 12). Аналогично звуковые колебания, воспринимаемые нами при слушании симфонического оркестра, представляют собой сумму колебаний воздуха, вызываемых каждым из музы-

---

[2] Приведем для справки значения A для нескольких показателей степени n: $n\to 0$, $A\to 2/\pi\approx 0{,}64$; $n=1$, $A=0{,}5$; $n=2$, $A=4/3\pi\approx 0{,}42$; $n=3$, $A=3/8=0{,}37$

кальных инструментов в отдельности, — и обязанности дирижера состоят в согласовании всей этой массы колебаний, для того чтобы эти колебания вместе звучали красиво. (Эффект чисто случайного наложения многих звуковых колебаний наблюдал каждый, кто слышал, как музыканты перед концертом настраивают свои инструменты, не обращая внимания один на другого.) Существует и множество (иногда очень важных) случаев, когда приходится иметь дело с накладывающимися одно на другое гармоническими колебаниями; более того, почти любое движение можно представлять себе как возникающее в результате сложения множества гармонических колебаний (см. § 9).

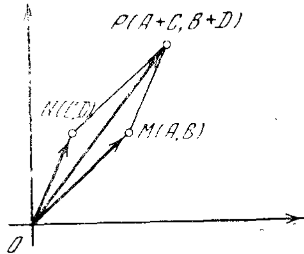


Рис. 10.7.1

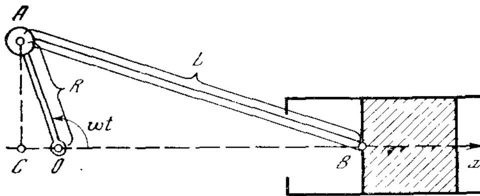


Рис. 10.7.2

Мы начнем со сложения д в у х гармонических колебаний. Наиболее простым здесь является случай, когда эти колебания имеют о д н у и т у ж е частоту (например, когда речь идет о совместном звучании двух играющих в унисон скрипок). В этом случае в результате мы получаем гармоническое колебание той же самой частоты. В самом деле, если два колебательных процесса описываются формулами

$$x_1 = A\cos\omega t + B\sin\omega t,$$
$$x_2 = C\cos\omega t + D\sin\omega t, \qquad (10.7.1)$$

то сумма $x = x_1 + x_2$ имеет вид

$$x = (A + C)\cos\omega t + (B + D)\sin\omega t. \qquad (10.7.2)$$

Связь колебания (2) с колебаниями (1) удобнее всего описать так: если колебаниям (1) сопоставить точки M и N с координатами (A, B) и (C, D) (или два вектора $\overrightarrow{OM}$ и $\overrightarrow{ON}$), то колебанию (2) будет отвечать вершина P параллелограмма $OMPN$ (или вектор $\overrightarrow{OP} = \overrightarrow{OM} + \overrightarrow{ON}$; см. рис. 1). При этом энергия колебаний (1) задается величинами $(OM)^2 = A^2 + B^2$ и $(ON)^2 = C^2 + D^2$ (ср. (4.2) и (2.8а)); энергия «составного» колебания (2) пропорциональна $(OP)^2$. Если начальные фазы колебаний (1) одинаковы, то отрезки OM и ON имеют одно и то же направление и длина отрезка OP равна с у м м е длин отрезков OM и ON; при этом энергия колебания (2) оказывается значительно больше суммарной энергии колебаний (1), ибо энергия задается к в а д р а т о м амплитуды колебаний, т. е. отрезка OM, соответственно ON и OP: так, если колебания (1) одинаковы, то энергия их суммы в четыре раза больше энергии каждого колебания. Если начальные фазы колебания (1) противоположны (т. е. отличаются на π), то отрезки OM и ON направлены в противоположные стороны — здесь амплитуда OP составного колебания равна р а з н о с т и амплитуд OM и ON составляющих колебаний. При этом колебания могут и полностью погасить друг друга, если их амплитуды (или энергии) одинаковы; в случае же несовпадающих и близких энергий колебаний x_1 и x_2, фазы которых расходятся на π, результирующее колебание будет иметь весьма малую энергию (см. упр. 1).

Сложнее обстоит дело в случае сложения колебаний р а з л и ч н о й частоты, когда составляющее колебание уже не является гармоническим. Для примера рассмотрим так называемый *кривошипно-шатунный механизм*, преобразующий поступательное движение во вращательное и наоборот; этот механизм составляет часть паровой машины. На рис. 2 кривошип OA связан с ползуном B (с поршнем, двигающимся взад и вперед внутри замкнутого цилиндра) с помощью шатуна AB; длину кривошипа OA мы обозначим через R, а длину шатуна AB — через L. Предположим, что кривошип OA равномерно вращается вокруг оси O с частотой ω, так что проекция C точки A на направле-

ние OB оси цилиндра, в котором ходит ползун B, совершает гармонические колебания вокруг точки O. При этом ползун B также будет совершать колебания вокруг какой-то «средней» точки — но эти колебания уже не будут гармоническими.

В самом деле, если $\widehat{AOB}=\omega t$, то $OC=R\cos\omega t$, $AC=R\sin\omega t$ и, значит,

$$BC=\sqrt{L^2-R^2\sin^2\omega t}=$$
$$=L\sqrt{1-\left(\frac{R}{L}\right)^2\sin^2\omega t}.$$

Таким образом, получаем

$$x=OB=R\cos\omega t+$$
$$+L\sqrt{1-\left(\frac{R}{L}\right)^2\sin^2\omega t}. \qquad (10.7.3)$$

Но в реальных механизмах обычно бывает $R\ll L$, так что отношение $\frac{R}{L}$ невелико (часто оно имеет величину порядка 1/5, так что $\left(\frac{R}{L}\right)^2\approx\frac{1}{25}$); при небольшом же z величину $\sqrt{1-z}$ можно заменить на $1-\frac{z}{2}$ (см. § 6.4). Поэтому сравнительно сложное выражение (3) можно упростить так:

$$x\approx R\cos\omega t+L\left(1-\frac{R^2}{2L^2}\sin^2\omega t\right)=$$
$$=R\cos\omega t+L\left[1-\frac{R^2}{2L^2}\frac{(1-\cos 2\omega t)}{2}\right]=$$
$$=\left(L-\frac{R^2}{4L}\right)+R\cos\omega t+$$
$$+\frac{R^2}{4L}\cos 2\omega t. \qquad (10.7.4)$$

Первое слагаемое $L-\frac{R^2}{4L}$ указывает среднее положение ползуна B, вокруг которого он колеблется, участвуя сразу в двух гармонических колебаниях: в колебании с частотой ω и амплитудой R, задаваемом движением точки C, и в колебании с вдвое большей частотой 2ω и амплитудой $\frac{R^2}{4L}$, связанном с изменением высоты CA точки A (рис. 3).

Своеобразная картина получается при сложении колебаний с **близкими** частотами ω_1 и ω_2. Если, например, $x_1=A\cos\omega_1 t$, $x_2=A\cos\omega_2 t$ (случай сложения колебаний одной амплитуды A), то

$$x=x_1+x_2=A\cos\omega_1 t+A\cos\omega_2 t=$$
$$=A(\cos\omega_1 t+\cos\omega_2 t)=$$
$$=A\cdot 2\cos\frac{\omega_1-\omega_2}{2}t\cos\frac{\omega_1+\omega_2}{2}t. \qquad \textbf{(10.7.5)}$$

При этом, очевидно, множитель $2A\cos\frac{\omega_1-\omega_2}{2}t$ меняется весьма мед-

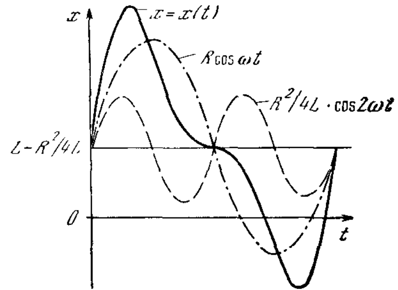


Рис. 10.7.3

ленно, а множитель $\cos\frac{\omega_1+\omega_2}{2}t$ соответствует колебаниям с частотой $\frac{\omega_1+\omega_2}{2}\approx\omega$. Колебания (5) называют *биениями* (ср. рис. 4, *а*); их часто несколько неточно описывают как гармонические колебания

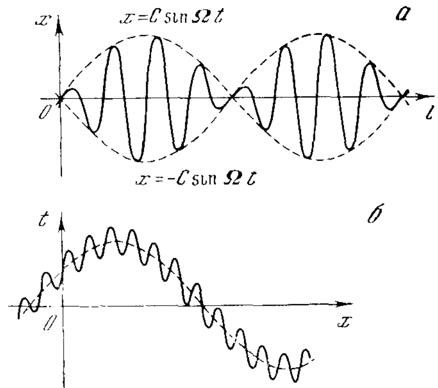


Рис. 10.7.4

с частотой $\omega\left(\approx\frac{\omega_1+\omega_2}{2}\right)$ и медленно меняющейся амплитудой $A(t)=2A\times\times\cos\frac{\omega_1-\omega_2}{2}t(=C\cos\Omega t)$. Аналогичная картина получается и в том случае, если фазы колебаний x_1 и x_2 различаются между собой — только в этом случае также и фаза суммарного колебания x испытывает медленные колебания (см. упр. 2а).

Напротив, если частоты ω и Ω составляющих колебаний **сильно разнятся** между собой, скажем если

$x_1 = A\cos\omega t$ и $x_2 = B\cos\Omega t$, где $\Omega \gg \omega$, то суммарное колебание $x = x_1 + x_2$ можно представлять себе как гармоническое колебание амплитуды B с частотой Ω вокруг медленно меняющегося положения равновесия $a = a(t) = A\cos\omega t$ (см. рис. 4, *б*, где еще $A \gg B$).

С биениями мы встречаемся в явлении океанских приливов и отливов: здесь уровень воды в океане меняется под действием притяжения Луны и притяжения Солнца, причем в связи с вращением Земли вокруг оси уровень воды меняет свое значение с периодом в 12 ч. Однако в связи с вращением Луны вокруг Земли период колебаний уровня воды, вызванных притяжением Луны, несколько отличается от периода колебаний, вызванного притяжением Солнца: если период «солнечных» приливов и отливов равен в точности 12 ч, то (более значительное) колебание уровня, вызванное притяжением Луны, имеет период, близкий к 12 $^3/_4$ ч. Так как амплитуды этих двух колебаний неодинаковы, то амплитуда сложного колебания никогда не равна нулю, т. е. оно не совпадает в точности с биениями, изображенными на рис. 4, *а*; однако здесь также можно говорить о медленном изменении амплитуды колебаний уровня воды, причем наибольшая высота прилива почти в 3 раза превосходит наименьшую (ср. упр. 2б). В радиотехнике также используются колебания с переменной амплитудой: так называемые «короткие» радиоволны с длиной волны порядка 20 м имеют частоту $1{,}5\cdot 10^7$ Гц (или 15 МГц), в то время как амплитуда этих колебаний меняется со звуковой частотой, т. е. имеет порядок сотен или нескольких тысяч колебаний в секунду (и лишь эти колебания воспринимает наше ухо).

Упражнения

10.7.1. Чему равны: а) энергия; б) амплитуда колебания, полученного сложением двух колебаний x_1 и x_2 с одинаковой частотой: $x_1 = A_1\cos(\omega t + \varphi_1)$ и $x_2 = A_2\cos(\omega t + \varphi_2)$?

10.2.2. а) Что представляет собой сумма гармонических колебаний $x_1 = a\cos(\omega_1 t + \varphi_1)$ и $x_2 = a\cos(\omega_2 t + \varphi_2)$, где ω_1 и ω_2 близки одна другой? б) Ответьте на тот же вопрос, когда колебания x_1 и x_2 имеют разные амплитуды a_1 и a_2.

§ 8. Задача о колеблющейся струне

В предыдущем параграфе мы рассматривали вопрос о функциях $x = f(t)$, представимых в виде суммы (суперпозиции) двух гармонических колебаний (с одинаковыми или разными частотами). Далее мы коснемся очень важной темы о представлении в виде суммы гармонических колебаний любых (непрерывных или даже не непрерывных) функций. К задаче о таком представлении функций математики и естествоиспытатели пришли, разбирая целый ряд естественно возникающих задач. Одной из первых подобных задач была известная *задача о колеблющейся струне.*

Рассмотрим расположенную в плоскости x, y струну, закрепленную в точках $x=0$ и $x=l$; здесь под струной понимается тонкая нить, которая может свободно изгибаться и вдоль которой действует постоянная сила натяжения. Выведем струну из состояния равновесия, при котором она совпадает с осью x, т. е. придадим ей произвольную форму $y = f(x)$ (рис. 1). Затем отпустим струну; нас интересует вопрос: как будет меняться ее форма под действием сил натяжения. При этом мы ограничимся случаем небольших деформаций струны, т. е. будем считать, что невелики как y (струна мало отклоняется от оси x), так и $\frac{\partial y}{\partial x}$ (деформированная струна по направлению все еще близка к оси x, т. е. является пологой, так что углы $\alpha = \operatorname{arc\,tg}\frac{\partial y}{\partial x}$, образованные касательной к струне с осью x, достаточно малы). Мы пишем $\frac{\partial y}{\partial x}$, а не $\frac{dy}{dx}$, потому что отклонение y струны меняется с течением времени, т. е. $y = y(x, t)$ представляет собой функцию двух переменных: абсциссы x и времени t. Выведем теперь дифференциальное уравнение, которому удовлетворяет величина $y(x, t)$.

Рассмотрим малый участок MN струны, ограниченный точками $M(x, y)$ и $N(x+dx, y+dy)$. Длина ds этого участка струны, очевидно, равна

$$ds = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + \left(\frac{\partial y}{\partial x}\right)^2} \approx$$

$$\approx dx\left[1 + \frac{1}{2}\left(\frac{\partial y}{\partial x}\right)^2\right] \approx dx$$

(ср. § 7.8 и 6.4): ведь мы считаем производную $\frac{\partial y}{\partial x}$ малой и пренебрегаем величинами порядка $\left(\frac{\partial y}{\partial x}\right)^2$. Так как мы рассматриваем колебания струны лишь в перпендикулярном оси x направлении, то скорость v точки M равна $\frac{\partial y}{\partial t}$, а ее ускорение a совпадает со второй (частной) производной $\frac{\partial^2 y}{\partial t^2}$; масса же участка MN струны равна $\sigma ds \approx \sigma dx$, где $\sigma = \text{const}$ — (линейная) плотность струны (струну мы считаем однородной). Уравнение движения участка MN струны мы получим, приравняв произведение $ma = \sigma dx \frac{\partial^2 y}{\partial t^2}$ действующей на этот участок силе.

Единственные рассматриваемые нами силы — это силы натяжения струны (другими словами, мы изучаем свободные колебания струны без воздействия каких бы то ни было внешних сил). На концы M и N участка MN струны действуют ведущие в сторону от MN силы R натяжения струны, направленные по касательным к струне. Составляющие этих сил, перпендикулярные оси x и вызывающие поперечное колебание струны, соответственно равны $R\sin\alpha$ и $R\sin(\alpha+d\alpha)$, где α и $\alpha+d\alpha$ — углы, образованные с осью x касательными к струне в точках $M\,(x, y)$ и $N\,(x+dx,\ y+dy)$ (см. рис. 1). Равнодействующая рассматриваемых сил равна

$$R\sin(\alpha+d\alpha) - R\sin\alpha \approx Rd(\sin\alpha)$$

(ср. § 4.1). Но мы знаем, что $\operatorname{tg}\alpha = \frac{\partial y}{\partial x}$ и что если угол α мал, то $\sin\alpha \approx \alpha$ и $\operatorname{tg}\alpha \approx \alpha$, так что можно написать

$$\sin\alpha \approx \operatorname{tg}\alpha \approx \frac{\partial y}{\partial x}$$

(здесь мы по-прежнему пренебрегаем величинами порядка α^2, или $\left(\frac{\partial y}{\partial x}\right)^2$). Поэтому

$$d\sin\alpha \approx d\left(\frac{\partial y}{\partial x}\right) = \frac{\partial^2 y}{\partial x^2}\,dx:$$

ведь время t в наших рассмотрениях пока считается закрепленным.

Таким образом, сила, вызывающая колебания участка MN струны, близка к $R\frac{\partial^2 y}{\partial x^2}dx$ — и уравнение колебаний струны можно записать так:

$$\sigma\frac{\partial^2 y}{\partial t^2}dx = R\frac{\partial^2 y}{\partial x^2}dx, \quad \text{или}$$

$$\frac{\partial^2 y}{\partial t^2} = c^2\frac{\partial^2 y}{\partial x^2}, \tag{10.8.1}$$

где $c^2 = \frac{R}{\sigma}$ $\left(\text{ясно, что } \frac{R}{\sigma} > 0\right)$. При этом решение уравнения (1) должно удовлетворять двум *граничным условиям*

$$y(0, t) = 0 \text{ при всех } t,\ y(l, t) = 0 \text{ при всех } t \tag{10.8.2а}$$

(концы струны закреплены) и двум *начальным условиям*

$$y(x, 0) = f(x),\quad \frac{\partial y(x, 0)}{\partial t} = 0 \quad \text{при всех } x \tag{10.8.2б}$$

(последнее условие означает, что в начальный момент времени $t=0$ струна

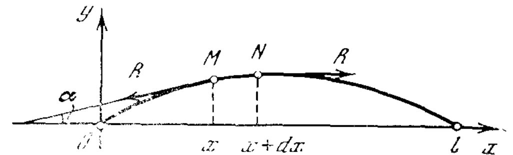


Рис. 10.8.1

никак не движется: мы оттянули струну, придав ей форму линии $y=f(x)$, и затем опустили ее [1]).

Уравнение (1) задачи о колеблющейся струне было впервые выписано и решено в 1747 г. Ж. Л. Д'Аламбером (см. упр. 3). Но большее значение имело несколько более позднее решение той же задачи, найденное Даниилом Бернулли (1753 г.). Бернулли рассуждал следующим образом. Будем искать частное решение уравнения (1), удовлетворяющее тем же граничным условиям (2а) (а также и второму начальному условию $\left.\frac{\partial y}{\partial t}\right|_{t=0}=0$) и имеющее форму произведения $y(x, t) = X(x)\,T(t)$ функции X, зависящей только от x, на функцию T, зависящую только от t. Подставляя в (1) значение $y = XT$, получим

$$XT'' = c^2X''T, \quad \text{или} \quad \frac{1}{c^2}\frac{T''}{T} = \frac{X''}{X}, \tag{10.8.3}$$

[1] Вместо (2б) можно было бы потребовать, чтобы скорость $v(x, 0) = \left.\frac{\partial y}{\partial t}\right|_{t=0}$ струны при всех x в начальный момент времени $t=0$ имела заданные значения: $v(x, 0) = g(x)$, где $g(x)$ известно (*общая* задача о закрепленной струне; см. упр. 2).

где, разумеется, $X''=\frac{d^2X}{dx^2}$ и $T''=\frac{d^2T}{dt^2}$ (ведь и X, и T — функции одной переменной!). Но так как $\frac{1}{c^2}\frac{T''}{T}$ зависит только от t, а $\frac{X''}{X}$ — только от x, то равенство этих величин означает независимость отношений $\frac{T''}{T}$ и $\frac{X''}{X}$ и от t, и от x:

$$\frac{X''}{X}=\frac{1}{c^2}\frac{T''}{T}=\text{const}. \qquad (10.8.4)$$

Ясно, что если отношения (4) **р а в н ы н у л ю**, то $X''\equiv 0$, т. е. $X=ax+b$ — **л и н е й н а я** функция переменного x. Но так как в силу (2а) должно быть $X(0)=X(l)=0$ (ведь $y=X(x)\,T(t)$, где, конечно, $T(t)\not\equiv 0$, ибо иначе наше решение ни малейшего интереса не представляло бы!), то равенство $X(x)=ax+b$ означает, что $b=0$ (так как $X(0)=0$) и $al=0$, т. е. $a=0$ (поскольку $X(l)=0$). Таким образом, никакого содержательного (т. е. не равного тождественно нулю) решения уравнения (1) мы здесь не получим.

Если отношения (4) **п о л о ж и т е л ь н ы**, то их можно обозначить через k^2:

$$\frac{X''}{X}=k^2, \quad \text{т. е.} \quad X''=k^2X. \qquad (10.8.5)$$

Уравнение (5) во многом аналогично уравнению гармонических колебаний (2.1). Так же как и в случае уравнения (2.1), частные его решения легко могут быть угаданы — таковыми являются функции $X_1=e^{kx}$ и $X_2=e^{-kx}$:

$$X_1'=ke^{kx} \quad \text{и} \quad X_1''=k^2e^{kx}=k^2X_1;$$

$$X_2'=-ke^{-kx}, \quad X_2''=(-k)^2e^{-kx}=k^2X_2$$

(ср. § 13.10). Произвольная комбинация

$$X=aX_1+bX_2=ae^{kx}+be^{-kx}$$

двух найденных решений также является (общим!) решением уравнения (5), позволяющим удовлетворить любым начальным условиям: $x_0=X(0)=A$, $X_0'=X'(0)=B$ (почему?). Однако и здесь желание удовлетворить условиям (2а) приводит к равенствам

$$X(0)=a+b=0,$$

$$X(l)=ae^{kl}+be^{-kl}=0,$$

т. е. $b=-a$ и $a(e^{kl}-e^{-kl})=0$,

удовлетворяющимся лишь при $a=b=0$. Таким образом, также и предположение о положительности стоящих в левой части (4) отношений не приводит ни к какому решению уравнения (1), отличному от нулевого решения: $X(x)\equiv 0$, а значит, и $y(x,t)\equiv 0$.

Наконец, пусть рассматриваемые отношения **о т р и ц а т е л ь н ы**: $\frac{X''}{X}=\frac{1}{c^2}\frac{T''}{T}=-k^2<0$, т. е.

$$\frac{X''}{X}=-k^2, \quad X''=-k^2X, \qquad (10.8.6а)$$

$$\frac{1}{c^2}\frac{T''}{T}=-k^2, \quad T''=-c^2k^2T. \qquad (10.8.6б)$$

В противоположность предшествующим двум случаям здесь уже возникают интересные решения уравнения (1).

Ясно, что (6а) и (6б) суть уравнения (2.1) гармонических колебаний для функций $X=X(x)$ и $T=T(t)$ (от (2.1) они отличаются только обозначениями!). Из (6а) вытекает, что

$$X=A\cos kx+B\sin kx \qquad (10.8.7)$$

(ср. § 2). А так как в силу первого из граничных условий (2а) $X(0)=0$, то коэффициент A в правой части (7) равен нулю: $X=B\sin kt$. Подставив теперь это значение X во второе граничное условие $X(l)=0$, найдем, что должно иметь место равенство $B\sin(kl)=0$, что при $B\neq 0$ (а случай $A=B=0$, конечно же, совершенно неинтересен!) возможно, лишь если

$$kl=n\pi, \quad k=\frac{\pi}{l}n, \quad n \text{ — целое число.} \qquad (10.8.8)$$

Таким образом, окончательно решение (7) уравнения (6а) принимает вид

$$X=B\sin\left(\frac{n\pi}{l}x\right), \qquad (10.8.9)$$

где B произвольно, а n — какое угодно целое число.

Аналогично второе уравнение гармонических колебаний (6б) (где теперь надо требовать выполнения условий (8)!) допускает решение

$$T=C\cos(ckt)+D\sin(ckt)=$$
$$=C\cos\left(\frac{cn\pi}{l}t\right)+D\sin\left(\frac{cn\pi}{l}t\right). \qquad (10.8.10)$$

А поскольку

$$T'=\frac{dT}{dt}=-\frac{cn\pi}{l}C\sin\left(\frac{cn\pi}{l}t\right)+$$
$$+\frac{cn\pi}{l}D\cos\left(\frac{cn\pi}{l}t\right),$$

то (предположенное выполняющимся) второе из начальных условий (2б) $\left.\frac{\partial y}{\partial t}\right|_{t=0}=0$, т. е. $T'(0)=0$, влечет за собой равенство $D=0$. Таким образом,

$$T=C\cos\left(\frac{cn\pi}{l}t\right)$$

и, значит,

$$y=XT=b_n\sin\left(\frac{n\pi}{l}x\right)\cos\left(\frac{cn\pi}{l}t\right), \tag{10.8.11}$$

где $b_n=BC$ — произвольное число, а (выбранное заранее) n — какое-то натуральное число [2].

Итак, мы нашли целое семейство (11) решений уравнения (1): ведь натуральное число n в (11) может быть каким угодно. Все эти решения имеют характер гармонических колебаний: каждая фиксированная точка струны, которой отвечает закрепленное значение $x=x_0$, колеблется по простому закону

$$y(x_0)=d_n\cos(\omega_n t) \tag{10.8.12}$$

с произвольной амплитудой $d_n=d_n(x_0)=b_n\sin\left(\frac{n\pi}{l}x_0\right)$, но со строго фиксированной частотой $\omega_n=\left(\frac{cn}{l}\right)\pi$. Это значение ω_n частоты зависит не только от физических характеристик струны (длины l, силы натяжения R и линейной плотности σ, отношения $\frac{R}{\sigma}=c^2$), но также и от (произвольно фиксированного) натурального числа n: различные допустимые частоты

$$\omega=\omega_1=\frac{c\pi}{l},\quad \omega_2=\frac{2c\pi}{l}=2\omega,$$
$$\omega_3=3\omega,\ \ldots \tag{10.8.13}$$

образуют арифметическую прогрессию — все они кратны минимальной частоте ω. При этом закон (11) колебания струны утверждает, что все ее участки колеблются «в унисон»: частоты ω_n для всех x одинаковы, а различаются колебания отдельных участков струны лишь (плавно меняющимися, можно сказать, согласованными) амплитудами $d_n=d_n(x)=b_n\sin\left(\frac{n\pi}{l}x\right)$.

Итак, мы нашли решение (11) (точнее, много решений (11)) уравнения (1), удовлетворяющее как граничным условиям (2а), так и второму из начальных условий (2б). Ну, а как обстоят дела с первым начальным условием? Ясно, что этому условию нам, вообще говоря, удовлетворить не удастся: хотя мы и располагаем еще постоянной b_n, которую можем выбрать как угодно, все

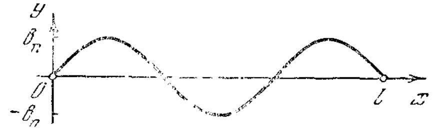


Рис. 10.8.2

равно очевидно, что начальная форма струны

$$y(x,\ 0)=b_n\sin\left(\frac{n\pi}{l}x\right)$$

никак не является произвольной— она имеет синусоидальный характер (с таким полупериодом, что l является его целым кратным; см. рис. 2, где $n=3$).

Как же нам поступить, если исходная функция $y(x, 0)=f(x)$, задающая первоначальную форму струны, не сводится к синусоиде? Гениальная идея Д. Бернулли (которую в XIX в. Ж. Б. Фурье развил в общий мощный метод решения разнообразных задач математической физики; см. § 9) состояла в сведении произвольного (казалось бы, достаточно сложно зависящего и от x, и от t) колебания струны к системе отдельных гармонических колебаний («маятников») (11). А именно, он предложил использовать *принцип суперпозиции*, т. е. тот факт, что любая сумма решений уравнения (1) также удовлетворяет этому уравнению.

Образуем теперь сумму

$$y=b_1\sin\left(\frac{\pi x}{l}\right)\cos\left(\frac{c\pi t}{l}\right)+$$
$$+b_2\sin\left(\frac{2\pi x}{l}\right)\cos\left(\frac{2c\pi t}{l}\right)+\ldots \tag{10.8.14}$$

частных решений (11) уравнения (1). Ясно, что эта сумма также удовлетворяет уравнению (1), обоим граничным условиям (2а) и второму из начальных условий (2б), поскольку этому уравне-

[2] Ранее мы считали число n целым, а не только натуральным (целым положительным). Однако ясно, что замена знака n в (11) приводит лишь к изменению знака b_n и не дает нам новых решений (1), а значение $n=0$ отвечает (неинтересному) нулевому решению $y=0$ уравнения.

нию и условиям удовлетворяет каждое слагаемое суммы.

Подставляя в (14) значение $t=0$ и учитывая первое начальное условие (2б), получаем

$$y(x,\ 0)=f(x)=b_1\sin\left(\frac{\pi x}{l}\right)+$$
$$+b_2\sin\left(\frac{2\pi x}{l}\right)+b_3\sin\left(\frac{3\pi x}{l}\right)+\ \ldots \qquad (10.8.15)$$

— и наша задача сводится к тому, чтобы определить из (15) коэффициенты b_1, b_2, b_3, ... представления (14) решения уравнения (1). Достижение этой цели вполне можно расценить как выдающийся успех — это означало бы, что

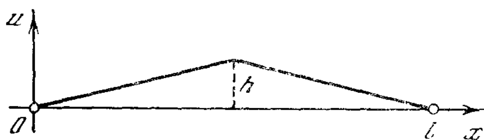


Рис. 10.8.3

«непрерывное» колебание $y=y(x,\ t)$ струны нам удалось сконструировать из дискретного (разъединенного) набора (11) отдельных «маятников» (гармонических колебаний), а общее уравнение (1) с помощью своеобразного «метода разделения переменных» (аргументов) x и t удалось свести к обыкновенным (не содержащим частных производных) и простым (решающимся в элементарных функциях) уравнениям (6а) и (6б). (Разумеется, представление (14) решения уравнения (1) может быть полезно только в том случае, если оно позволяет, ограничившись суммой некоторого (желательно — небольшого) числа первых членов стоящей в правой части (14) суммы, получить достаточно полное представление об искомой функции $y=y(x,\ t)$.)

Бернулли много занимался теорией колебаний и был хорошо знаком с процессом разложения произвольного колебания на отдельные гармоники (т. е. на простые гармонические колебания): $f(t)=\sum\limits_k c_k\cos(\omega_k t+\varphi_k)$. Исходя из своего опыта, он полагал, что всякую функцию $f(x)$ можно представить в виде суммы гармоник (15), т. е. что формула (14) дает *общее решение* уравнения колебания струны (1) (коэффициенты b_1, b_2, b_3, ... разложения (14) определяются из условия (15)). Это заключение вызвало ожесточенную критику знаменитого Л. Эйлера (см. с. 182). Эйлер исходил из того, что в случае простейшей начальной функции $f(x)$, изображенной на рис. 3, когда мы просто отводим струну в ее середине $x=\frac{l}{2}$ на фиксированную (небольшую) высоту h, а затем отпускаем струну, функция $f(x)$ на разных участках своего определения задается разными уравнениями:

$$f(x)=\begin{cases}\dfrac{2hx}{l} & \text{при } x\leqslant\dfrac{l}{2},\\ -\dfrac{2hx}{l}+2h & \text{при } x\geqslant\dfrac{l}{2},\end{cases} \qquad (10.8.16)$$

в то время как формула (15) задает единое выражение для $f(x)$ на всем протяжении $0\leqslant x\leqslant l$. Эйлера поддержал и Д'Аламбер, который также считал невозможным представление любой функции в виде (15) (якобы опровергающий Бернулли пример (16) впервые был указан именно Д'Аламбером). Однако Бернулли не сдавался: исходя из опыта механических явлений, он твердо стоял на возможности представления (15) для любой функции $y(x,0)$,— и этот спор сыграл очень большую роль как в прояснении точного смысла самого понятия «функция», так и в теории представлений функций так называемыми тригонометрическими рядами, т. е. суммами (15) гармоник (графиков гармонических колебаний; см. § 9).

Упражнения

10.8.1. Пусть $l=\pi$, $c=1$ и $f(x)=\sin^3 x$. Найдите решение (14) уравнения (1), удовлетворяющее условиям (2а) и (2б).

10.8.2. Как видоизменяется решение Д. Бернулли уравнения (1) в том случае, когда начальные условия решения уравнения (1) имеют вид: $y(x,\ 0)=f(x)$, $\left.\dfrac{\partial y(x,\ t)}{\partial t}\right|_{t=0}=g(x)$ (мы придали струне начальное положение $y(x,\ 0)=f(x)$ и начальные скорости $v(x,\ 0)=\left.\dfrac{\partial y(x,\ t)}{\partial t}\right|_{t=0}=g(x)$ и затем отпустили ее)?

10.8.3 (решение Д'Аламбера уравнения (1)). Докажите, что любая функция $y(x,\ t)=\varphi(x+ct)+\psi(x-ct)$, где φ и ψ — произвольные функции одной переменной, удовлетворяет уравнению (1). Как надо подобрать φ и ψ для того, чтобы решение $y(x,\ t)$ удовлетворяло граничным условиям (2а) и начальным

условиям (2б)? Граничным условиям (2а) и начальным условиям: $y(x, 0)=f(x)$, $v(x, 0)=g(x)$ (ср. упр. 2)?

§ 9. Гармонический анализ функций. Ряды Фурье

Пусть мы имеем произвольную функцию $y=f(x)$, заданную на каком-то конечном интервале; нам будет удобно считать, что этот интервал определяется неравенством $-\pi \leqslant x \leqslant \pi$ (см., впрочем, с. 337 и упр. 2). Задача состоит в том, чтобы найти коэффициенты разложения функции $f(x)$ в т р и г о н о м е т р и ч е с к и й р я д, т. е. представление ее в виде суммы гармоник 1, $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, $\cos 3x$, $\sin 3x$, . . . и т. д., взятых с какими-то численными коэффициентами. Для этого напишем

$$f(x)=\frac{a_0}{2}+a_1\cos x+b_1\sin x+ + a_2\cos 2x+b_2\sin 2x+a_3\cos 3x+ + b_3\sin 3x+\ldots, \qquad (10.9.1)$$

где числа $\frac{a_0}{2}$, a_1, b_1, a_2, b_2, . . . мы пока считаем неизвестными (не зависящее от x слагаемое суммы (1) нам будет удобно записать как $\frac{a_0}{2}$, а не как a_0).

Отметим теперь свойство о р т о г о н а л ь н о с т и тригонометрических функций:

$$\int_{-\pi}^{\pi}\cos mx\cos nx dx=\int_{-\pi}^{\pi}\sin mx\sin nx dx=0$$

при любых натуральных m и n, где $m\neq n$;

$$\int_{-\pi}^{\pi}\cos mx\sin nx dx=0$$

при любых (натуральных) m и n; (10.9.2)

$$\int_{-\pi}^{\pi}\cos^2 mx dx=\int_{-\pi}^{\pi}\sin^2 mx dx=\pi$$

при всех (натуральных) m.

Эти формулы вытекают из того, что, как известно,

$$\cos mx\cos nx={}^1/_2[\cos(m-n)x+ + \cos(m+n)x],$$

$$\sin mx\sin nx={}^1/_2[\cos(m-n)x- - \cos(m+n)x], \qquad (10.9.3)$$

$$\cos mx\sin nx={}^1/_2[\sin(m+n)x- - \sin(m-n)x];$$

частными случаями первых двух соотношений (3), отвечающими равенству $m=n$, являются известные зависимости

$$\cos^2 mx={}^1/_2(1+\cos 2mx),$$
$$\sin^2 mx={}^1/_2(1-\cos 2mx) \qquad (10.9.3а)$$

(ибо $\cos 0=1$). Далее, чтобы убедиться в справедливости (2), остается только проинтегрировать в пределах от $-\pi$ до π правые части всех формул (3) и (3а), воспользовавшись изложенными в гл. 5 правилами.

Пусть теперь мы имеем представление (1) функции $y=f(x)$. При этом функция $f(x)$ нам задана; требуется определить коэффициенты a_0, a_1, b_1, a_2, b_2, a_3, b_3, . . . тригонометрического ряда. Для того чтобы найти коэффициент a_m ряда, умножим левую и правую части равенства (1) на $\cos mx$ и проинтегрируем обе части в пределах от $-\pi$ до π. Слева мы тогда получим величину $\int_{-\pi}^{\pi} f(x)\cos mx dx$; справа же в силу (2) у нас выпадут все слагаемые, кроме слагаемого

$$\int_{-\pi}^{\pi} a_m\cos^2 mx dx= =a_m\int_{-\pi}^{\pi}\cos^2 mx dx=a_m\pi, \qquad (10.9.4)$$

поскольку остальные члены стоящей в правой части (1) суммы после интегрирования обратятся в

$$\int_{-\pi}^{\pi} a_n\cos nx\cos mx\, dx= =a_n\int_{-\pi}^{\pi}\cos nx\cos mx\, dx=0,$$
$$\int_{-\pi}^{\pi} b_n\sin nx\cos mx\, dx= =b_n\int_{-\pi}^{\pi}\sin nx\cos mx\, dx=0 \qquad (10.9.5)$$

(при $n=0$ первая из формул (5) обращается в равенство $\int\limits_{-\pi}^{\pi} \frac{a_0}{2} \cos mx\,dx = \frac{a_0}{2} \int\limits_{-\pi}^{\pi} \cos mx\,dx = 0$). Окончательно мы приходим к соотношению

$$\int\limits_{-\pi}^{\pi} f(x) \cos mx\,dx = a_m \pi,$$

или $a_m = \frac{1}{\pi} \int\limits_{-\pi}^{\pi} f(x) \cos mx\,dx$;

$m = 1, 2, 3, \ldots$ (10.9.6)

Совершенно аналогично, умножая обе части (1) на $\sin mx$ и интегрируя затем в пределах от $-\pi$ до π, получаем

$$\int\limits_{-\pi}^{\pi} f(x) \sin mx\,dx = b_m \pi,$$

или $b_m = \frac{1}{\pi} \int\limits_{-\pi}^{\pi} f(x) \sin mx\,dx$;

$m = 1, 2, 3, \ldots$ (10.9.6a)

Наконец, просто интегрируя обе части (1) в пределах от $-\pi$ до π и воспользовавшись тем, что

$$\int\limits_{-\pi}^{\pi} \cos nx\,dx = \int\limits_{-\pi}^{\pi} \sin nx\,dx = 0 \text{ и}$$

$$\int\limits_{-\pi}^{\pi} dx = 2\pi,$$

получаем

$$\int\limits_{-\pi}^{\pi} f(x)\,dx = \int\limits_{-\pi}^{\pi} \frac{a_0}{2}\,dx = \frac{a_0}{2} \cdot 2\pi,$$

или $a_0 = \frac{1}{\pi} \int\limits_{-\pi}^{\pi} f(x)\,dx$ (10.9.7)

— частный случай формулы (6), отвечающий значению $m = 0$.[1]

Формулы для коэффициентов тригонометрического ряда упрощаются в том случае, когда рассматриваемая функция является четной или нечетной. Если функция $f(x)$ — четная, т. е. $f(-x) = f(x)$, то

$$\int\limits_{-\pi}^{\pi} f(x) \sin mx\,dx = \int\limits_{-\pi}^{0} f(x) \sin mx\,dx + \int\limits_{0}^{\pi} f(x) \sin mx\,dx =$$

$$= \int\limits_{-\pi}^{0} f(-x) \sin(-mx)\,d(-x) + \int\limits_{0}^{\pi} f(x) \sin mx\,dx = \int\limits_{\pi}^{0} f(x) \sin mx\,dx + \int\limits_{0}^{\pi} f(x) \sin mx\,dx = 0,$$

ибо $f(-x) = f(x)$ и $\sin(-mx) = -\sin x$, $d(-x) = -dx$. Поэтому в нашем случае

$$f(x) = \frac{a_0}{2} + a_1 \cos x + a_2 \cos 2x + a_3 \cos 3x + \ldots, \qquad (10.9.8)$$

где

$$a_0 = \frac{1}{\pi} \int\limits_{-\pi}^{\pi} f(x)\,dx = \frac{2}{\pi} \int\limits_{0}^{\pi} f(x)\,dx;$$

$$a_m = \frac{1}{\pi} \int\limits_{-\pi}^{\pi} f(x) \cos mx\,dx = \frac{1}{\pi} \left(\int\limits_{-\pi}^{0} f(x) \cos mx\,dx + \int\limits_{0}^{\pi} f(x) \cos mx\,dx \right) = \frac{2}{\pi} \int\limits_{0}^{\pi} f(x) \cos mx\,dx. \qquad (10.9.9)$$

Точно так же, если функция $f(x)$ — нечетная, т. е. $f(-x) = -f(x)$, то

$$a_m = 0 \text{ и } b_m = \frac{2}{\pi} \int\limits_{0}^{\pi} f(x) \sin mx\,dx \qquad (10.9.10)$$

(докажите это сами!), т. е.

$$f(x) = b_1 \sin x + b_2 \sin 2x + b_3 \sin 3x + \ldots \qquad (10.9.11)$$

Таким образом, четные функции раскладываются в ряд (1) по одним лишь косинусам, а нечетные — только по синусам.

[1] Именно для того, чтобы прийти к формуле (7), являющейся частным случаем (6), мы писали в правой части (1) $\frac{a_0}{2}$, а не a_0.

Пример 1. Пусть

$$f(x)=\begin{cases}\pi+x & \text{при} \quad -\pi\leqslant x\leqslant 0,\\ \pi-x & \text{при} \quad 0\leqslant x\leqslant\pi\end{cases}$$

(рис. 1, *а*; ср. с рис. 8.3). Эта функция четная, поэтому она раскладывается в ряд по одним только косинусам. При этом

$$a_0=\frac{2}{\pi}\int_0^\pi f(x)\,dx=\frac{2}{\pi}\int_0^\pi(\pi-x)\,dx=$$

$$=\frac{2}{\pi}\left(\pi x-\frac{x^2}{2}\right)\Big|_0^\pi=\frac{2}{\pi}\left(\pi^2-\frac{\pi^2}{2}\right)=\pi,$$

$$a_m=\frac{2}{\pi}\int_0^\pi(\pi-x)\cos mx\,dx=$$

$$=\frac{2}{\pi}\left(\pi\int_0^\pi\cos mx\,dx-\int_0^\pi x\cos mx\,dx\right)=$$

$$=\frac{2}{\pi}\left[\pi\int_0^\pi\cos mx\,dx-\frac{1}{m}\int_0^\pi x\,d(\sin mx)\right]=$$

$$=\frac{2}{\pi}\left[\pi\int_0^\pi\cos mx\,dx-\frac{1}{m}(x\sin mx)\Big|_0^\pi+\right.$$

$$\left.+\frac{1}{m}\int_0^\pi\sin mx\,dx\right]=\frac{2}{\pi}\left(\frac{\pi}{m}\sin mx\Big|_0^\pi-\right.$$

$$\left.-\frac{x}{m}\sin mx\Big|_0^\pi-\frac{1}{m^2}\cos mx\Big|_0^\pi\right)=$$

$$=\begin{cases}0 & \text{при } m \text{ четном},\\ \dfrac{4}{\pi m^2} & \text{при } m \text{ нечетном}.\end{cases}$$

Окончательно мы получаем

$$f(x)=\frac{\pi}{2}+\frac{4}{\pi}\cos x+\frac{4}{9\pi}\cos 3x+$$

$$+\frac{4}{25\pi}\cos 5x+\ldots \qquad (10.9.12)$$

На рис. 1, *а* последовательно изображены функции $\varphi_0(x)=\frac{\pi}{2}$ (мелкий пунктир); $\varphi_1(x)=\frac{\pi}{2}+\frac{4}{\pi}\cos x$ (пунктир); $\varphi_2(x)=\frac{\pi}{2}+\frac{4}{\pi}\cos x+\frac{4}{9\pi}\cos 3x$ (тонкая сплошная линия).

Пример 2. Пусть

$$g(x)=\begin{cases}1 & \text{при} \quad 0<x\leqslant\pi,\\ -1 & \text{при} \quad -\pi\leqslant x<0\end{cases}$$

(рис. 1, *б*). Эта функция нечетная, поэтому она разлагается в ряд по одним синусам. При этом

$$b_m=\frac{2}{\pi}\int_0^\pi\sin mx\,dx=\frac{2}{m\pi}(-\cos mx)\Big|_0^\pi=$$

$$=\begin{cases}\dfrac{4}{m\pi} & \text{при } m \text{ нечетном},\\ 0 & \text{при } m \text{ четном},\end{cases}$$

так что

$$g(x)=\frac{4}{\pi}\left(\sin x+\frac{\sin 3x}{3}+\frac{\sin 5x}{5}+\ldots\right). \qquad (10.9.13)$$

На рис. 1, *б* последовательно изображены функции $\psi_1(x)=\frac{4}{\pi}\sin x$ (мелкий

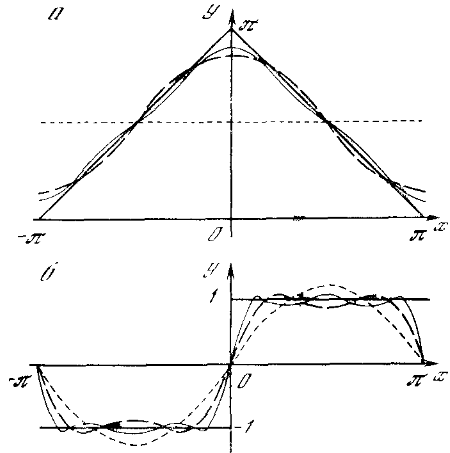


Рис. 10.9.1

пунктир); $\psi_2(x)=\frac{4}{\pi}\left(\sin x+\frac{\sin 3x}{3}\right)$ (пунктир); $\psi_3(x)=\frac{4}{\pi}\left(\sin x+\frac{\sin 3x}{3}+\frac{\sin 5x}{5}\right)$ (тонкая сплошная линия).

До сих пор мы говорили лишь о функциях, заданных на интервале $-\pi\leqslant x\leqslant\pi$. Но почти ничем практически не отличается и случай функции $f(x)$, заданной на произвольном интервале $-l\leqslant x\leqslant l$: ведь ясно, что область задания подобной функции мы всегда можем «подобно раздуть», сведя ее к интервалу от $-\pi$ до π (ср. с § 1.7).

Введем новую переменную $x_1=\left(\frac{\pi}{l}\right)x$, или $x=\left(\frac{l}{\pi}\right)x_1$; тогда $f(x)$ можно будет также записать как $f\left(\frac{l}{\pi}x_1\right)=f_1(x_1)$, где функция f_1 аргумента $x_1=\left(\frac{\pi}{l}\right)x$ уже определена при $-\pi\leqslant x_1\leqslant\pi$. При этом разложение (1), в котором только аргу-

мент x и функция f заменены на x_1 и f_1, порождает следующее разложение исходной функции $y=f(x)$:

$$y=\frac{a_0}{2}+a_1\cos\left(\frac{\pi}{l}x\right)+b_1\sin\left(\frac{\pi}{l}x\right)+$$
$$+a_2\cos\left(\frac{2\pi}{l}x\right)+b_2\sin\left(\frac{2\pi}{l}x\right)+$$
$$+a_3\cos\left(\frac{3\pi}{l}x\right)+b_3\sin\left(\frac{3\pi}{l}x\right)+\ldots, \qquad (10.9.14)$$

причем здесь

$$a_k=\frac{1}{l}\int_{-l}^{l}f(x)\cos\left(\frac{k\pi x}{l}\right)dx,$$
$$b_k=\frac{1}{l}\int_{-l}^{l}f(x)\sin\left(\frac{k\pi x}{l}\right)dx \qquad (10.9.15)$$

(см. упр. 1).

Укажем еще, что функцию $y=f(x)$, заданную, скажем, в интервале $0\leqslant$

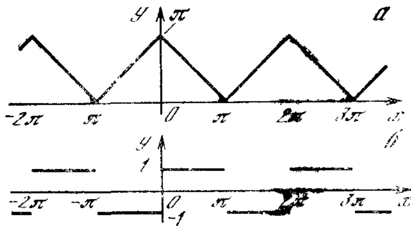


Рис. 10.9.2

$\leqslant x\leqslant l$, всегда можно продолжить на отрицательные значения x «по четности», положив $f(-x)=f(x)$, что приводит к следующему представлению $f(x)$:

$$f(x)=\frac{a_0}{2}+a_1\cos\left(\frac{\pi}{l}x\right)+a_2\cos\left(\frac{2\pi}{l}x\right)+$$
$$+a_3\cos\left(\frac{3\pi}{l}x\right)+\ldots \qquad (10.9.16)$$

где, очевидно,

$$a_k=\frac{2}{l}\int_{0}^{l}f(x)\cos\left(\frac{k\pi x}{l}\right)dx;$$
$$k=0,\ 1,\ 2,\ \ldots \qquad (10.9.17)$$

(почему? см. упр. 1). Можно также задать значения $f(x)$ при отрицательных x так, чтобы функция стала нечетной, т. е. положить $f(-x)=-f(x)$; это приводит к разложению $f(x)$ по синусам:

$$f(x)=b_1\sin\left(\frac{\pi}{l}x\right)+b_2\sin\left(\frac{2\pi}{l}x\right)+$$
$$+b_3\sin\left(\frac{3\pi}{l}x\right)+\ldots, \qquad (10.9.18)$$

где

$$b_k=\frac{2}{l}\int_{0}^{l}f(x)\sin\left(\frac{k\pi x}{l}\right)dx;$$
$$k=1,\ 2,\ 3,\ \ldots \qquad (10.9.19)$$

В задаче о колебании струны мы встретились именно с последним выражением для функции $y(x,0)=f(x)$.

Заметим еще, что правую часть равенства (1) можно, разумеется, понимать как функцию, определенную не только в интервале $-\pi\leqslant x\leqslant\pi$, но на всей числовой оси. Ведь все стоящие в правой части (1) тригонометрические функции $\cos x$, $\sin x$, $\cos 2x$, $\sin 2x$, $\cos 3x$, $\sin 3x$, ... являются *периодическими* с наименьшим общим периодом 2π: если $\varphi(x)$ — какая угодно из рассматриваемых функций, то $\varphi(x+2k\pi)=\varphi(x)$ при любом целом k. Поэтому и сумма стоящего в правой части (1) ряда (мы будем записывать эту сумму по-прежнему как $f(x)$) является периодической с периодом 2π:

$$f(x+2k\pi)=f(x) \text{ при всех целых } k,$$

ибо все слагаемые правой части (1) при замене x на $x+2k\pi$ сохраняют свое значение. Аналогично этому правая часть более общей формулы (14) задает периодическую функцию с периодом $2l$. Соответственно этому про формулы (1) и (14) иногда говорят, что они доставляют разложение в тригонометрический ряд периодических функций с периодом 2π (соответственно $2l$). Так, на рис. 2 изображены («полные») функции $f(x)$ и $g(x)$, определяемые разложениями (12) и (13).

Ясно, что все фигурирующие в этом параграфе представления (1), (8), (11), (14), (16), (18) функций в виде бесконечных сумм тригонометрических функций (в виде тригонометрических рядов, как мы неоднократно говорили выше) представляют ценность и интерес лишь в том случае, когда соответствующие ряды сходятся, если употреблять здесь терминологию § 6.3, т. е. когда сумма конечного (желательно даже — небольшого) числа первых членов, скажем, ряда (1)

$$s_n=s_n(x)=\frac{a_0}{2}+a_1\cos x+b_1\sin x+$$
$$+a_2\cos 2x+b_2\sin 2x+\ldots$$
$$\ldots+a_n\cos nx+b_n\sin nx \qquad (10.9.20)$$

дает достаточно хорошее представление о правой части (1), а при $n \to \infty$ сумма s_n стремится к $f(x)$, так что приближенное равенство

$$f(x) \approx s_n(x) \qquad (10.9.21)$$

является тем более точным, чем больше n. Но в историческом споре Д. Бернулли со знаменитым Л. Эйлером и Ж. Л. Д'Аламбером прав был именно Бернулли: оказалось, что практически любую[2] (периодическую или заданную на конечном интервале) функцию можно разложить в тригонометрический ряд! Это утверждение означает, что, скажем, каждая заданная на интервале $-\pi \leqslant x \leqslant \pi$ функция $f(x)$ (или, что то же самое, каждая периодическая функция $f(x)$ с периодом 2π) может быть представлена в виде (1), т. е. для нее имеет место приближенное равенство (21), где правая часть (21) имеет вид (20). При этом в точках, где функция $f(x)$ непрерывна, сумма $s_n(x)$ будет близка к $f(x)$ (причем, разумеется, тем ближе к $f(x)$, чем больше n), а в точках разрыва $f(x)$, где эта функция совершает скачок, так что ее значение слева $f(x-0)$ отлично от значения справа $f(x+0)$, сумма $s_n(x)$ с ростом n приближается к среднему арифметическому $^1/_2\,[f(x-0)+f(x+0)]$ этих значений[3]. При этом скорость сходимости сумм $s_n(x)$ к функции $f(x)$ существенно зависит от характера функции: она является наибольшей в случае гладких функций; наличие разрывов у производной функции (см. выше пример 1) эту скорость существенно снижает, а наличие разрывов у самой функции (как в случае примера 2) снижает ее еще значительнее, что вынуждает в этом случае для представления функции с заданной точностью использовать больше членов тригонометрического ряда.

Поясним, не претендуя при этом ни на какую строгость, причины, обусловливающие влияние «степени гладкости» функции $f(x)$ на скорость сходимости представляющего ее тригонометрического ряда. Ясно, что, для того чтобы, скажем, в представлении (1) можно было отбросить все члены стоящей в правой части (1) бесконечной суммы, кроме известного числа начальных ее членов, надо, чтобы отброшенные члены быстро убывали, чтобы влияние даже бесконечного их числа оказалось сравнительно небольшим. А так как сами по себе тригонометрические функции ограничены (ни одна из них не превосходит по абсолютной величине единицу!), то надо, чтобы при движении вдоль ряда быстро убывали задаваемые формулами (6) и (6а) коэффициенты ряда. Но для гладких функций убывание этих коэффициентов убедительно следует из колебательного (знакопеременного!) характера стоящих под интегралом в формулах для a_m и b_m функций $\cos mx$ и $\sin mx$.

В самом деле, при большом m период, скажем, функции $\sin(mx)$ оказывается достаточно малым, так что (гладкая!) функция $f(x)$ на протяжении этого периода почти не успевает измениться. А отсюда следует, что отвечающая одному периоду функции $\sin mx$ часть выражающего коэффициент b_m интеграла $\int f(x)\sin mx\,dx$ будет весьма близка к нулю, поскольку $f(x)$ здесь можно считать постоянной и значения рассматриваемого интеграла, отвечающие двум полупериодам функции $\sin mx$, на которых она имеет разные знаки, почти полностью сократятся.

Более точный расчет показывает, что также и сумма всех тех частей стоящих в правых частях (6) и (6а) интегралов, которые отвечают отдельным волнам косинусоиды $y=\cos mx$ и синусоиды $y=\sin mx$, при больших m оказывается весьма малой, что и обеспечивает малость коэффициентов ряда (1), — а тем самым и сходимость этого ряда. При этом в случае негладкости функции $f(x)$, когда в какой-то точке она резко меняет свое направление (см. рис. 2, *а*), убывание коэффициентов ряда оказывается не столь быстрым, как в тех случаях, когда функция всюду является гладкой; еще хуже обстоит дело, если сама функция претерпевает разрыв (ср. рис. 2, *б*), в сколь угодно малой окрестности которого $f(x)$ никак нельзя считать постоянной величиной! И в самом деле, мы видим, что коэффициенты b_m отвечающего разрывной функции $g(x)$ ряда (13) с ростом m убывают (всего лишь) как $\frac{1}{m}$, в то время как в случае непрерывной функции $f(x)$ примера 1 коэффициенты a_m ряда (12) убывают, грубо

---

[2] Для того чтобы представление функции $f(x)$ в виде тригонометрического ряда было невозможно, надо, чтобы $f(x)$ имела на интервале, на котором она задана (или на протяжении одного своего периода), бесконечно много разрывов («скачков») или бесконечно много максимумов и минимумов!

[3] Так, для рассмотренной в примере 2 функции надо считать, что $g(0-0)=-1$, $g(0+0)=1$; суммы $s_n(x)$ при $x=0$ сходятся здесь к значению $\frac{1}{2}[(-1)+1]=0$.

говоря, как $\frac{1}{n^2}$, а для гладкой функции убывание коэффициентов отвечающего ей тригонометрического ряда оказывается еще более быстрым.

Вернемся теперь к знаменитой дискуссии между Д'Аламбером, Эйлером и Д. Бернулли по вопросу о характере колебания закрепленной в двух концах струны. Частные решения (8.11) рассматриваемой задачи были получены еще в 1713—1715 гг. упоминавшимся уже в этой книге Тейлором, который, впрочем, неосновательно полагал, что никакие другие решения здесь невозможны. Достижение Д. Бернулли состояло в том, что он считал возможным смешение (или, как говорят сегодня, наложение) отдельных колебаний (8.11), — полагал, что струна может звучать, издавая одновременно тоны различной частоты. При этом *основной тон* струны задается главной (или первой) ее частотой ω; дополнительные же частоты 2ω, 3ω и т. д. (см. (8.13)) отвечают так называемым *обертонам*[4]. Именно богатство физического содержания решения задач о колебании струны, даваемого представлением (8.13) функции $y(x, t)$, а также полное совпадение получаемых таким путем результатов с физическими экспериментами заставляло Бернулли так настойчиво держаться за это решение и отвергать критику глубоко им уважаемых Эйлера (являвшегося его другом) и Д'Аламбера.

Дальнейшее усовершенствование методики Бернулли связано в первую очередь с исследованиями Ж. Б. Фурье. Фурье впервые изложил в связном виде общую теорию разложения функций в тригонометрические ряды, основанную на полученных выше простых формулах для определения коэффициентов ряда; он дал также множество примеров разложения конкретных функций. Но еще важнее были глубокие применения подобных разложений к конкретным задачам математической физики, например к задаче о распространении тепла. Эти приложения базировались на той же идее, что и у Бернулли: начальные условия соответствующей задачи, задающиеся произвольной функцией $f(x)$, Фурье первоначально считал «синусоидальными», т. е. ограничивался функциями типа $\sin mx$ или $\cos mx$. Для этих простых начальных условий задачу решить оказывалось сравнительно несложно, а общее решение получалось отсюда с помощью принципа суперпозиции и возможности разложения произвольной функции $f(x)$ в тригонометрический ряд. И не случайно, несмотря на то, что отдельные примеры разложений функций в тригонометрические ряды до Фурье рассматривались многими учеными (скажем, Л. Эйлером, Д. Бернулли, К. Ф. Гауссом и др.), все такие ряды сегодня принято называть *рядами Фурье*.

Процедура разложения функции $f(x)$ в ряд Фурье (или в тригонометрический ряд) носит название *спектрального анализа* функции. Так, спектральное разложение (12) фигурирующей в примере 1 функции $f(x)$ показывает, что эта функция состоит из постоянной $\frac{\pi}{2}$ и гармоник, причем эти гармоники $\cos x$, $\cos 3x$, $\cos 5x$, ... берутся с коэффициентами $\frac{4}{\pi}, \frac{4}{9\pi}, \frac{4}{25\pi}, \dots$, а функция $g(x)$ примера 2 раскладывается в сумму гармоник $\frac{4}{\pi}\sin x, \frac{4}{3\pi}\sin 3x, \frac{4}{5\pi}\sin 5x, \dots$ (см. (13)).

Спектральное разложение (периодических) функций играет очень большую роль в самой математике и ее приложениях (в частности, в изучении всякого рода колебательных процессов); оно зачастую производится автоматически, при помощи специальных механизмов — *(гармонических) анализаторов*. Само выражение «спектральное разложение» (или «спектральный анализ») функций связано с тем, что собственно такой характер имеет разложение света в совокупность волн (гармонических колебаний) разной длины волны (или разного периода), причем для световых волн каждой длине волны (или периоду) отвечает свой цвет. В общем случае спектральное разложение произвольной (непериодической!) функции содержит «гармо-

[4] Соотношения между основным тоном и обертонами звучащей струны, сводящиеся к простым отношениям длин периодов отвечающих им синусоидальных волн, и связь этих целочисленных отношений с благозвучием издаваемых струнами звуковых колебаний рассматривались еще древнегреческими учеными пифагорейской школы (V в. до н. э.). Полученные здесь пифагорейцами результаты сыграли очень важную роль в сложившемся у них убеждении в простоте и познаваемости мира, причем ключом к познанию природы пифагорейцы считали математику.

ники» $a(\omega)\cos(\omega t+\varphi(\omega))$, отвечающие **всевозможным** частотам ω, так что суммы Фурье приходится заменять *интегралами Фурье*, распространенными по широкому «спектру» частот ω (случай «непрерывного спектра» функции). Мы, однако, не коснемся здесь этой более сложной постановки вопроса[5].

Упражнения

10.9.1. Докажите формулы (10), (15), (17), (19).

10.9.2. Разложите в тригонометрический ряд функцию $y=f(x)$, заданную в каком угодно интервале $a\leqslant x\leqslant b$. [Указание. Подстановка $x=\frac{b-a}{2\pi}x_1+\frac{a+b}{2}$ переводит функцию $f(x)$ в функцию $f_1(x_1)$, заданную на интервале $-\pi\leqslant x_1\leqslant\pi$.]

10.9.3. Разложите в тригонометрические ряды функции: а) $y=x^2$, $-\pi\leqslant x\leqslant\pi$;

б) $y=\begin{cases}c_1x & \text{при } -l\leqslant x\leqslant 0,\\ c_2x & \text{при } 0\leqslant x\leqslant l;\end{cases}$

в) $y=\begin{cases}\sin x & \text{при } 0\leqslant x\leqslant \pi,\\ 0 & \text{при } -\pi\leqslant x\leqslant 0.\end{cases}$

---

[5] По этому поводу см., например, книгу [15].

Глава 11

ТЕПЛОВОЕ ДВИЖЕНИЕ МОЛЕКУЛ. РАСПРЕДЕЛЕНИЕ ПЛОТНОСТИ ВОЗДУХА В АТМОСФЕРЕ

§ 1. Условие равновесия в атмосфере

Рассмотрим вопрос о законе распределения плотности воздуха в атмосфере по высоте. Хорошо известно, что на большой высоте воздух менее плотен, давление воздуха меньше, чем на уровне моря. Причина зависимости давления от высоты очевидна. Выделим мысленно цилиндрический объем высоты Δh; если площадь его основания равна S, то объем будет равен $S\Delta h$. Воздух, находящийся в этом объеме (средняя плотность $\bar{\rho}$, масса $m=\bar{\rho}S\Delta h$), притягивается к Земле, т. е. испытывает силу тяготения, направленную вниз и равную $mg=\bar{\rho}Sg\Delta h$. Однако этот воздух не опускается вниз, а находится в покое, ибо на него снизу (на высоте h от поверхности Земли) действует давление $p(h)$, большее, чем давление $p(h+\Delta h)$ сверху, на высоте $h+\Delta h$ (рис. 1). Сила давления на нижнее основание цилиндра равна $Sp(h)$; она уравновешивает сумму силы давления на верхнее основание и силы тяжести:

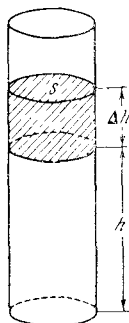


Рис. 11.1.1

$$Sp(h)=Sp(h+\Delta h)+\bar{\rho}Sg\Delta h. \qquad (11.1.1)$$

Формулу (1) можно переписать так:

$$p(h)-p(h+\Delta h)=\bar{\rho}g\Delta h. \qquad (11.1.2)$$

Будем считать, что Δh весьма мало. Тогда вместо средней плотности $\bar{\rho}$ можно просто говорить о плотности ρ, так как высоты h и $h+\Delta h$ весьма близки и средняя плотность $\bar{\rho}$ весьма мало отличается от плотности $\rho(h)$ на высоте h. Поэтому, считая, что в левой части (2) стоит величина $-dp$, и заменяя Δh на dh, получим

$$\frac{dp}{dh}=-g\rho. \qquad (11.1.3)$$

Тем самым мы получили дифференциальное уравнение для зависимости $p=p(h)$ давления от высоты, в которое вошла также и плотность воздуха ρ.

Мы будем предполагать, что температура атмосферы одинакова на всех высотах. На самом деле температура воздуха зависит от притока тепла от солнца и отвода тепла в основном за счет излучения его в мировое пространство воздухом или, точнее, содержащимися в воздухе водяными парами и углекислотой. Небольшая часть лучей солнечного спектра поглощается верхними, очень разреженными слоями воздуха. Большая часть энергии солнечного света доходит до поверхности Земли, поглощается почвой, а уже от почвы нагревается воздух. Поэтому в действительности в атмосфере получается довольно сложное распределение температуры: на уровне земли температура, как мы хорошо знаем, колеблется примерно от —40 до 40° С в зависимости от географического положения и времени года; на высоте около 15 км температура минимальна (около —80° С) и примерно одинакова зимой и летом на всем земном шаре; на больших высотах температура возрастает, достигая 60—75° С на высоте 50—60 км.

Измерения, проделанные в последнее время при помощи искусственных спутников Земли, показывают, что на высоте 300—1000 км плотность воздуха хотя и мала, но все же значительно больше, чем предполагалось раньше. Как мы увидим ниже, сравнительно большая плотность воздуха указывает на то, что в этих слоях температура воздуха весьма высока. К тому же значительная часть молекул кислорода и азота распадается на этой высоте на атомы, ионы и электроны.

Если бы не было притока и отвода тепла, т. е. если рассматривать тепло-

изолированный столб воздуха, то с течением времени температура во всем столбе должна была бы выровняться. Ниже мы будем рассматривать именно такой идеализированный случай полного равновесия, теплового и механического. Тепловое равновесие заключается в том, что температура везде одинакова, следовательно, нет потоков тепла (если бы в различных точках воздушного столба температура была различна, то от более нагретых точек тепло текло бы к менее нагретым, т. е. возникли бы потоки тепла). Механическое равновесие заключается в том, что равнодействующая всех сил на любой объем воздуха, выделенный в атмосфере, равна нулю; при этом надо рассматривать силу тяжести воздуха в объеме и силу давления на всю поверхность, ограничивающую данный объем.

При том распределении давления, которое удовлетворяет уравнению (3), атмосфера может находиться в состоянии покоя.

Так как мы рассматриваем высоты h, малые по сравнению с радиусом Земли, то ускорение свободного падения g можно считать постоянной величиной.

§ 2. Связь между плотностью и давлением

В уравнении (1.3) фигурируют две неизвестные нам величины: давление p и плотность воздуха ρ; поэтому прежде всего надо установить связь между ними.

По закону Бойля—Мариотта *при постоянной температуре t произведение давления газа p на занимаемый им объем V постоянно*: для данной массы m_0 газа и при данной температуре $pV=a$, где a — постоянная величина. Если плотность газа равна ρ, то $m_0=V\rho$. Следовательно, $V=\frac{m_0}{\rho}$, а так как $p=\frac{a}{V}$, то

$$p=b\rho, \qquad (11.2.1)$$

где $b=\frac{a}{m_0}$. Таким образом, *давление газа прямо пропорционально его плотности*.

Коэффициент пропорциональности для воздуха комнатной температуры легко найти. Известно, что атмосферное давление на уровне моря p_0 приблизительно равно 10^5 Н/м² (=10^5 Па). Плотность воздуха ρ_0 при давлении p_0 равна[1] примерно 1,3 кг/м³. Подставляя эти значения p_0 и ρ_0 в (1), получим $p_0=b\rho_0$, откуда[2]

$$b\approx 10^5/1{,}3\approx 7{,}7\cdot 10^4 \text{ м}^2/\text{с}^2.$$

Для дальнейшего нам нужно не только численное значение b для воздуха комнатной температуры, но и общее выражение постоянной b для любого газа и при любой температуре. Для этого воспользуемся законом Клапейрона:

$$pV=RT, \qquad (11.2.2)$$

где V — объем, занимаемый одним молем (одной грамм-молекулой) газа; T — абсолютная температура (отсчитываемая от абсолютного нуля, соответствующего $-273°$ С[3]); R — так называемая *универсальная газовая постоянная*. Известно, что при температуре 0° С (=273 К) при атмосферном давлении, т. е. при $p_0=10^5$ Па, один моль газа занимает объем, примерно равный 22,4 л = $2{,}24\cdot10^{-2}$ м³ (закон Авогадро). Отсюда $10^5\cdot 2{,}24\cdot 10^{-2}=273\,R$, т. е.

$$R\approx 8{,}3\ \ [(\text{Н/м}^2)\cdot\text{м}^3/(\text{моль}\cdot\text{град})=\text{Дж}/(\text{моль}\cdot\text{град})].$$

Обозначим относительную молекулярную массу газа через M. Для водорода $M=2$, для гелия $M=4$, для азота $M=28$, для воздуха среднее значение $M=29{,}4$. По определению, 1 моль газа объемом V содержит M г, или $M\cdot10^{-3}$ кг вещества. Значит, плотность ρ связана с V соотношением

$$\rho=10^{-3}\frac{M}{V}, \quad \text{или} \quad V=10^{-3}\frac{M}{\rho}.$$

---

[1] Эту величину легко найти экспериментально при помощи взвешивания. Берут сосуд известного объема с краном и сначала взвешивают его с воздухом, а затем без воздуха, откачав его из сосуда вакуумным насосом.

[2] Величина b имеет размерность квадрата скорости. В действительности эта величина тесно связана со скоростью молекул и со скоростью звука: квадрат скорости звука примерно равен $1{,}4\,b$ (выводить это соотношение мы не будем).

[3] Абсолютную температуру обозначают буквой К по имени английского ученого Кельвина; например, 20° С=293 К читается: 20 градусов Цельсия равны 293 градусам Кельвина (принято не употреблять значок градуса, когда речь идет об абсолютной температуре, оставляя только букву К).

Здесь плотность ρ выражена в кг/м^3, V — в м^3. Подставляя это выражение для V в (2), получим

$$p=10^3\rho\frac{RT}{M}. \qquad (11.2.3)$$

Сравнивая формулы (3) и (1), находим

$$b=10^3\frac{RT}{M}. \qquad (11.2.4)$$

Наконец, выразим давление через число молекул n, содержащихся в единице объема газа. Известно, что 1 моль любого вещества содержит около $6\cdot10^{23}$ молекул. Величина $6\cdot10^{23}$ называется *числом Авогадро* и обозначается A. Таким образом, масса одной молекулы (в кг)

$$m=\frac{M}{A}\cdot10^{-3}\approx\frac{1}{6\cdot10^{26}}M. \qquad (11.2.5)$$

Если 1 моль газа занимает объем V, то число молекул в единице объема n есть $n=\frac{A}{V}$. Плотность газа $\rho=nm$. Закон Клапейрона (2) дает

$$p=n\frac{RT}{A}=nkT,$$

где k — так называемая *постоянная Больцмана*:

$$k=\frac{R}{A}\approx\frac{8,3}{6\cdot10^{23}}=1,38\cdot10^{-23}\ \frac{\text{Дж}}{\text{град}}.$$

Величина R относится к условно выбранному количеству вещества — одному молю, поэтому в размерность R входит «моль». Величина k относится к одной молекуле, поэтому k имеет размерность Дж/град. Величина kT имеет размерность энергии (Дж). В § 4 будет показано, что в атмосфере величина kT равна средней потенциальной энергии одной молекулы в поле тяжести при температуре T. Средняя кинетическая энергия поступательного движения одной молекулы равна $\frac{3}{2}kT$.

§ 3. Распределение плотности

Из формулы (2.1) находим $\rho=\frac{p}{b}$. Подставляя это значение в дифференциальное уравнение (1.3) для давления, получим

$$\frac{dp}{dh}=-\frac{g}{b}p.$$

Решение этого уравнения имеет вид $p=Ce^{-\frac{g}{b}h}$, где C должно быть определено из начального условия. Пусть $p=p_0$ при $h=0$; тогда

$$p=p_0e^{-\frac{g}{b}h}. \qquad (11.3.1)$$

Разделив обе части (1) на b, получим

$$\rho=\rho_0e^{-\frac{g}{b}h}, \qquad (11.3.2)$$

где ρ_0 — плотность воздуха при $h=0$ (т. е. на уровне моря). Из формулы (1) видно, что на «характерной высоте» $H=\frac{b}{g}$ над уровнем моря давление воздуха уменьшается в e раз. Используя величину H, формулы (1) и (2) можно переписать так:

$$p=p_0e^{-h/H},\quad \rho=\rho_0e^{-h/H}. \qquad (11.3.3)$$

Подсчитаем величину H, пользуясь формулой $H=\frac{b}{g}$:

$$H\approx(7,7\cdot10^4)/10=7,7\times$$
$$\times10^3\ [(\text{см}^2/\text{с}^2):(\text{м}/\text{с}^2)=\text{м}]\approx7,7\ \text{км}$$

(это значение H отвечает средней температуре около 20° С).

С увеличением высоты в арифметической прогрессии давление и плотность уменьшаются в геометрической прогрессии: если $h=0$, то $p=p_0$, $\rho=\rho_0$; если $h=H$, то $p=\frac{p_0}{e}\approx0,368p_0$, $\rho\approx0,368\rho_0$; если $h=2H$, то $p=\frac{p_0}{e^2}\approx0,135p_0$, $\rho\approx\approx0,135\rho_0$; если $h=3H$, то $p=\frac{p_0}{e^3}\approx\approx0,05p_0$, $\rho\approx0,05\rho_0$ и т. д.

Получим теперь формулу, связывающую величину H с величиной kT. Мы знаем, что $H=\frac{b}{g}$; пользуясь формулами (2.4) и (2.5), находим

$$H=10^3\frac{RT}{Mg}=\frac{RT}{Amg},\quad\text{или}\quad H=\frac{kT}{mg}. \qquad (11.3.4)$$

Зная зависимость плотности от высоты, можно выразить через ρ_0 и H полную массу воздуха m_a в столбе с площадью основания 1 м^2. Действительно,

$$m_a=\int_0^\infty\rho\,dh=\int_0^\infty\rho_0e^{-h/H}dh.$$

Введем новую переменную $z=\frac{h}{H}$; тогда

$$dz=\frac{1}{H}dh,$$

$$m_a=\rho_0 H\int\limits_0^\infty e^{-z}dz=-\rho_0 He^{-z}\Big|_0^\infty=\rho_0 H.$$

При помощи соотношения $m_a=\rho_0 H$ подсчитаем (для контроля) еще раз величину H. Так как атмосферное давление равно $\approx 10^5$ Па ($=10^5$ Н/м^2), то масса воздуха в столбе с основанием 1 м^2 давит на землю с силой 10^5 Н, т. е. она примерно равна $\frac{10^5}{g}\approx 10^4$ кг. Таким образом, $m_a=10^4$ кг/м^2. Зная, что $\rho_0\approx\approx 1{,}3$ кг/м^3, получим отсюда

$$H=\frac{m_a}{\rho_0}=\frac{10^4}{1{,}3}\approx 7{,}7\cdot 10^3\text{ м}=7{,}7\text{ км}$$

в полном соответствии с прежним расчетом.

Найдем, наконец, среднюю высоту $\bar h$, на которой находится воздух, т. е. высоту центра тяжести вертикального цилиндрического столба воздуха. Чтобы не вводить лишних величин, рассматриваем столб воздуха с основанием 1 м^2; однако ясно, что высота центра тяжести от площади основания цилиндра не зависит. Приходящаяся на слой между значениями h и $h+dh$ высоты масса воздуха равна $dm=\rho dh$. Поэтому средняя высота равна

$$\bar h=\frac{\int\limits_0^\infty h\,dm}{m_a}=\frac{\int\limits_0^\infty h\rho(h)\,dh}{\int\limits_0^\infty \rho(h)\,dh}.$$

Вычислим интеграл, стоящий в числителе. Пользуясь второй из формул (3) и делая затем замену $\frac{h}{H}=z$, т. е. $dh=Hdz$, получаем

$$\int\limits_0^\infty h\rho(h)\,dh=\int\limits_0^\infty h\rho_0 e^{-h/H}dh=$$

$$=\rho_0 H^2\int\limits_0^\infty ze^{-z}dz=\rho_0 H^2$$

(ибо в силу формулы интегрирования по частям $\int\limits_0^\infty ze^{-z}dz=-ze^{-z}\Big|_0^\infty+\int\limits_0^\infty e^{-z}dz=(-ze^{-z}-e^{-z})\Big|_0^\infty=1$; ср. (7.5.19). Окончательно находим

$$\bar h=\frac{\rho_0 H^2}{\rho_0 H}=H. \tag{11.3.5}$$

Таким образом, та высота H, на которой плотность и давление воздуха убывают в e раз, одновременно является и средней высотой, на которой находится воздух.

Аналогичный результат получился раньше, при рассмотрении радиоактивного распада (§ 8.1): если вероятность распада равна ω, $\frac{dn}{dt}=-\omega n$, $n=n_0e^{-\omega t}$, то за время $\tau=\frac{1}{\omega}$ количество радиоактивного вещества убывает в e раз; при этом среднее время жизни $\bar t$ радиоактивного атома оказалось равным той же величине: $\bar t=\tau=\frac{1}{\omega}$.

Напомним, что простая зависимость (3) плотности и давления от высоты относится к случаю постоянной температуры. В действительности распределение плотности и давления несколько отличается от формулы (3) и зависит от времени года и т. п.

Упражнения

11.3.1. Найдите давление воздуха в шахте на глубине 1 км; 3 км; 10 км.

11.3.2. Найдите зависимость давления воздуха от высоты при температуре воздуха —40° С; 40° С.

11.3.3. Пусть температура воздуха меняется при изменении высоты по закону $\frac{dT}{dh}=-\alpha T_0$, где T_0 — температура воздуха на поверхности Земли, α — постоянный коэффициент. Найдите зависимость давления воздуха от высоты.

11.3.4. Известно, что в условиях упр. 3 величина $\alpha\approx 0{,}037\cdot 10^{-5}$ см$^{-1}$. Определите, пользуясь результатом упр. 3, давление воздуха в шахте на глубине 1 км; 3 км; 10 км, считая, что на поверхности Земли температура равна 0° С. Сравните результаты с результатами упр. 1.

§ 4. Молекулярно-кинетическая теория распределения плотности

В предыдущих параграфах было найдено распределение плотности воздуха по высоте при действии силы тяжести в состоянии равновесия. При этом мы рассматривали воздух как сплошную

среду с данной зависимостью давления от плотности.

Подойдем теперь к полученному результату с иной точки зрения, а именно с точки зрения молекулярной теории. Будем рассматривать отдельные молекулы и их движение. Представление о том, что вещество состоит из отдельных атомов, было высказано еще в Древней Греции. Однако движение молекул и связь его с теплотой впервые стал рассматривать великий русский ученый М. В. Ломоносов, который является, таким образом, основателем молекулярно-кинетической теории.

Газообразное состояние отличается от жидкого и твердого тем, что в газе молекулы можно рассматривать как независимые, не взаимодействующие между собой. Движение молекул газа представляет собой свободный полет по инерции. Время от времени молекулы газа сталкиваются между собой. В обычных условиях столкновения молекул между собой происходят чрезвычайно часто, участки пути, которые молекулы пролетают от одного столкновения до другого, весьма малы.

При атмосферном давлении и температуре 0° С в 22,4 л газа содержится 1 моль вещества, т. е. $6\cdot 10^{23}$ молекул; в 1 м^3 содержится $n\approx\frac{6\cdot 10^{23}}{2{,}24\cdot 10^{-2}}\approx\approx 2{,}7\cdot 10^{25}$ молекул.

Грубо будем рассматривать молекулы как шарики[1] радиусом около $2\cdot 10^{-10}$ м. Тогда, для того чтобы произошло столкновение двух молекул, надо, чтобы траектория центра одной молекулы попала в мишень радиусом $4\cdot 10^{-10}$ м вокруг центра другой молекулы. Площадь такой мишени $\sigma=\pi r^2\approx 5\cdot 10^{-19}$ м^2. Значит, на пути в 1 м данная молекула сталкивается со всеми теми молекулами, центры которых находятся в цилиндре с площадью основания $5\cdot 10^{-19}$ м^2 и длиной 1 м. Объем такого цилиндра (в м^3) равен σ и число молекул в нем равно $n\sigma$, где n — число молекул в 1 м^3.

Таким образом, на пути в 1 м молекула испытывает $n\sigma$ столкновений. Поэтому среднее расстояние свободного полета от одного столкновения до другого

$$l=\frac{1}{n\sigma}\approx 7{,}5\cdot 10^{-8}\ \text{м}.$$

Эта величина называется *длиной свободного пробега*.

Вследствие столкновений молекула летит не по прямой, а по ломаной линии. Однако объем цилиндра, образованного ломаными, практически не отличается от объема прямого цилиндра, так что наш расчет сохраняет силу и здесь.

На самом деле надо еще учесть, что движется не только рассматриваемая молекула, но и те молекулы, с которыми она сталкивается. Можно доказать, что это обстоятельство мало меняет длину свободного пробега, а именно уменьшает ее всего в 1,5 раза, так что далее мы будем считать, что $l\approx 5\cdot 10^{-8}$ м.

Скорости молекул v в диапазоне обычных жизненных температур имеют порядок величины 300—500 м/с. Следовательно, время свободного пробега, т. е. среднее время от одного столкновения до другого, будет порядка

$$\tau\approx\frac{5\cdot 10^{-8}}{4\cdot 10^{2}}\approx 10^{-10}\ \text{с}.$$

На первый взгляд величины $l\approx\approx 5\cdot 10^{-8}$ м и $\tau\approx 10^{-10}$ с очень малы. Однако их надо сравнить с размером молекулы радиуса $r\approx 2\cdot 10^{-10}$ м, и с длительностью самого столкновения, которая меньше $\frac{r}{v}$, т. е. имеет величину порядка 10^{-13} с. Из этого сравнения видно, что молекулы газа редко сталкиваются одна с другой: при атмосферном давлении молекулы 99,9 % времени проводят в полете и лишь 0,1 % времени находятся в состоянии столкновений.

Столкновения молекул между собой в газе не влияют на давление газа и не влияют на закон распределения плотности газа в атмосфере. Подтверждение этого заключается в законах Бойля—Мариотта и Клапейрона. В § 2 эти законы были записаны в виде $p=nkT$.

Давление газа зависит от числа молекул в единице объема, но радиус молекул r и их сечение σ не входят в формулу. Значит, величины r и σ не могут войти и в формулу распределения плотности по высоте.

[1] В действительности двухатомные молекулы, например молекулы кислорода или азота, уместнее представлять себе не как шарики, а как пары слепленных шаров вроде арахиса с двумя ядрами в одной скорлупе.

Перепишем формулу распределения плотности (3.2), выразив b через молекулярные величины. Так как $b = 10^3 \frac{RT}{M} = \frac{AkT}{Am} = \frac{kT}{m}$, то

$$\rho = \rho_0 e^{-\frac{gh}{b}} = \rho_0 e^{-\frac{mgh}{kT}} \qquad (11.4.1)$$

(ср. (3.3), (3.4)).

Разделим обе части формулы (1) на m, где m обозначает массу одной молекулы. Заметим, что $\frac{\rho}{m} = n$ — число молекул в единице объема на высоте h, $\frac{\rho_0}{m} = n_0$ — число молекул в единице объема на уровне моря. Формула (1) принимает теперь вид

$$n = n_0 e^{-\frac{mgh}{kT}}. \qquad (11.4.2)$$

Величина mgh есть потенциальная энергия молекулы массы m, находящейся на высоте h, если за нуль принять потенциальную энергию молекулы на уровне моря: ведь потенциальную энергию молекулы $u(0)$ на уровне моря можно выбрать произвольно (см. § 9.2). Тогда

$$u(h) = u(0) + mgh,$$

откуда

$$mgh = u(h) - u(0).$$

Формулу (2) можно теперь записать так:

$$n(h) = n(0)\, e^{-\frac{u(h)-u(0)}{kT}}.$$

Это — закон распределения числа молекул по высоте. Ему можно придать следующий вид:

$$n(h) = Be^{-\frac{u(h)}{kT}},$$

где B есть константа, определяемая по значению плотности на уровне моря ($h=0$):

$$n(0) = Be^{-\frac{u(0)}{kT}}.$$

Замечательно то обстоятельство, что плотность молекул на той или иной высоте зависит только от потенциальной энергии молекул в данном месте: масса молекулы m, ускорение свободного падения g и высота h вошли в формулу (2) как раз в том сочетании (mgh), в котором они входят в выражение для потенциальной энергии u.

Найдем **среднее значение** потенциальной энергии молекулы:

$$\bar{u} = \overline{mgh} = mg\bar{h} = mgH$$

(см. формулу (3.5)). Пользуясь формулой (3.4), получим

$$\bar{u} = mgH = mg\frac{kT}{mg} = kT.$$

Итак, средняя потенциальная энергия одной молекулы равна kT.

Мы установили, что распределение молекул воздуха в атмосфере зависит от температуры и потенциальной энергии молекул. Однако при данной средней потенциальной энергии $\bar{u}$, равной kT, разные молекулы имеют разную потенциальную энергию: данному значению $\bar{u}$ отвечает определенное *распределение* молекул по потенциальной энергии. Часть молекул, а именно те из них, которые находятся на высоте, меньшей H, имеет потенциальную энергию, меньшую, чем kT. Найдем отношение числа таких молекул к общему числу молекул. Это отношение равно

$$\int_0^H ndh \Big/ \int_0^\infty ndh = n_0 \int_0^H e^{-\frac{mgh}{kT}}dh \Big/ n_0 \int_0^\infty e^{-\frac{mgh}{kT}}dh.$$

Подсчитаем входящие сюда интегралы:

$$\int_0^H e^{-\frac{mgh}{kT}}dh = -\frac{kT}{mg}e^{-\frac{mgh}{kT}}\Big|_0^H = \frac{kT}{mg}\left(1 - e^{-\frac{mgH}{kT}}\right) = \frac{kT}{mg}(1 - e^{-1}),$$

$$\int_0^\infty e^{-\frac{mgh}{kT}}dh = -\frac{kT}{mg}e^{-\frac{mgh}{kT}}\Big|_0^\infty = \frac{kT}{mg}.$$

Поэтому

$$\int_0^H ndh \Big/ \int_0^\infty ndh = \frac{kT}{mg}(1 - e^{-1}) \Big/ \frac{kT}{mg} = 1 - e^{-1} \approx 0{,}63.$$

Таким образом, 63% всех молекул имеют потенциальную энергию меньше средней, а 37% имеют потенциальную энергию больше средней. При этом нетрудно подсчитать, что потенциальную энергию больше $2kT$ имеют 14% моле-

кул, а больше $3kT$ — только 5% всех молекул и т. д. Вообще доля молекул, имеющих потенциальную энергию, большую данной величины u, равна $e^{-u/kT}$. Этот вывод является очень общим — он относится и к кинетической энергии молекул, чему будет посвящен следующий параграф.

§ 5. Броуновское движение и распределение молекул по кинетической энергии

Более ста лет назад английский ботаник Браун (или Броун, как у нас принято писать эту фамилию) заметил, что взвешенные в воде или в иной жидкости мельчайшие видимые в микроскоп пылинки находятся в непрерывном движении. Эйнштейн высказал предположение, что это движение пылинок представляет собой их тепловое движение. Отсюда, в частности, следовал вывод, что пылинки не будут все лежать на

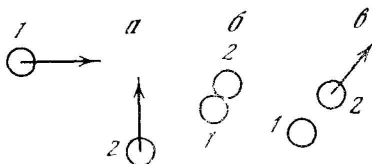


Рис. 11.5.1

дне сосуда, а будут распределены по высоте по такому же закону, по которому распределены молекулы.

Если пылинка имеет форму шара диаметром $d=5\cdot10^{-7}$ м, то ее объем равен $\frac{\pi d^3}{6}\approx 6{,}5\cdot10^{-20}$ м^3, и при плотности $\rho=2\cdot10^3$ кг/м^3 масса пылинки близка к $1{,}3\cdot10^{-16}$ кг. Необходимо, однако, учесть еще и закон Архимеда! Масса вытесненной пылинкой воды (плотность 10^3 кг/м^3) равна половине массы самой пылинки — и с учетом этого обстоятельства действующая на частицу сила тяжести отвечает половине ее массы, т. е. примерно $6{,}5\cdot10^{-17}$ кг. При комнатной температуре $T=17°$ С$=290$ К такие пылинки распределены по высоте (согласно формуле (4.2)) по закону

$$n\approx n_0\exp\left(-\frac{6{,}5\cdot10^{-17}\cdot9{,}8}{290\cdot1{,}38\cdot10^{-23}}h\right),$$

или $n=n_0e^{-1{,}6\cdot10^5 h}$. Таким образом, число пылинок в единице объема уменьшается в e раз при увеличении высоты на величину $\frac{1}{1{,}6\cdot10^5}$ м $\approx 0{,}62\cdot10^{-5}$ м.

Наблюдая распределение по высоте пылинок известного размера и плотности, можно получить значение постоянной Больцмана k. С другой стороны, закон Клапейрона дает величину $R=kA$, что позволяет найти число Авогадро. Эта работа, проведенная Эйнштейном и Перреном в 1903—1907 гг., является важнейшим экспериментальным подтверждением всей атомно-молекулярной теории и сыграла огромную роль в развитии физики.

При движении молекул под действием силы тяжести все время происходит превращение энергии: если молекула в данный момент движется вниз, то потенциальная энергия переходит в кинетическую; если же молекула движется вверх, то кинетическая энергия переходит в потенциальную. Когда газ находится в состоянии равновесия, т. е. давление газа уравновешивает силу тяжести, то в действительности молекулы газа движутся хаотически с большими скоростями. Однако, если мысленно выделить в газе какую-нибудь горизонтальную плоскость, то число молекул, проходящих за единицу времени через эту плоскость снизу вверх, будет равно числу молекул, проходящих за то же время сверху вниз, так что в среднем газ покоится. В состоянии равновесия переходы кинетической энергии в потенциальную и потенциальной энергии в кинетическую взаимно уравновешены, так как число молекул, двигающихся вверх, примерно равно числу молекул, двигающихся вниз.

Отметим, что при хаотическом движении отдельные (одинаковые) молекулы имеют различные скорости, т. е. различные кинетические энергии. Действительно, если сталкиваются между собой два шарика, скорости которых равны по величине, но направлены под углом одна к другой, то после столкновения скорости шариков могут стать различными. На рис. 1 показан пример столкновения, после которого один шарик, летящий слева, останавливается, а другой, двигающийся снизу, движется с удвоенной энергией (*а* — шарики до столкновения, *б* — в момент столкновения, *в* — после столкновения). Обратите внимание на то, как расположены шарики в момент столкновения: если бы второй шарик во время столкновения был расположен ниже первого, то он

остановился бы, отдав всю энергию первому шарику.

Так как при движении молекул происходит взаимное превращение кинетической и потенциальной энергий, то естественно предположить что распределение молекул по кинетической энергии похоже на их распределение по потенциальной энергии.

Приведем без доказательства результат расчетов, проделанных в конце XIX в. Максвеллом и Больцманом. Число молекул, имеющих составляющие скорости вдоль оси x между v_x и v_x+dv_x, вдоль оси y между v_y и v_y+dv_y и вдоль оси z между v_z и v_z+dv_z, равно

$$dn=\frac{n_0}{(2\pi kT/m)^{3/2}}\exp\left[-\frac{m\left(v_x^2+v_y^2+v_z^2\right)}{2kT}\right]\times \times dv_x dv_y dv_z, \qquad (11.5.1)$$

где n_0 — полное число молекул, m — масса одной молекулы. Отметим, что $v_x^2+v_y^2+v_z^2=v^2$, где v — скорость молекулы. Поэтому (1) в показателе степени содержит величину $\frac{mv^2}{2}:kT$, т. е. отношение кинетической энергии к потенциальной. Средняя кинетическая энергия, подсчитанная согласно закону (1), оказалась равной $\frac{3}{2}kT$. Для числа молекул n, кинетическая энергия которых больше заданной величины E, получилась довольно сложная зависимость. Однако эта сложная зависимость приближенно может быть описана простой формулой

$$n\approx n_0e^{-E/kT}. \qquad (11.5.2)$$

Из закона (2) получается неверное значение средней кинетической энергии молекул

$$\overline{E_{\text{кин}}}=\frac{\int_0^\infty n\,dE}{n_0}=\int_0^\infty e^{-E/kT}dE=kT$$

вместо $\frac{3}{2}kT$. Этот закон дает заметные отклонения от истинного, если E есть величина порядка kT. Однако при $E\gg kT$ расхождение между точным и приближенным законами становится несущественным.

Отметим, что при одинаковой температуре молекулы с различными массами имеют равные средние кинетические энергии и одинаково распределены по величине кинетической энергии, так как средняя скорость молекулы пропорциональна $\frac{1}{\sqrt{m}}$, где m — масса молекулы.

Рассматривая столкновения молекул со стенками сосуда, можно найти давление газа; оно оказывается равным

$$p=\frac{2}{3}n_0\bar{E}_{\text{кин}}.$$

Полагая здесь $\bar{E}_{\text{кин}}=\frac{3}{2}kT$, получим закон Клапейрона

$$p=n_0kT.$$

При столкновении молекул между собой происходит не только обмен кинетическими энергиями между молекулами, но и превращение кинетической энергии движения молекул в энергию вращения молекулы, а также в энергию колебаний атомов в молекуле, т. е. во внутреннюю энергию молекулы. Возможен и обратный процесс, когда при столкновении часть внутренней энергии молекул переходит в кинетическую энергию. Поэтому естественно, что и распределение молекул по величине их внутренней энергии W также подчиняется закону пропорциональности числа молекул с энергией W величине $e^{-W/kT}$: экспоненциальная зависимость числа частиц, обладающих той или иной энергией, от величины энергии является универсальным законом природы.

§ 6. Скорости химических реакций

Для чего нужен закон распределения молекул по кинетической энергии? Такие важные характеристики газа, как его давление на стенки сосуда, теплоемкость, общий запас энергии в объеме газа, являются средними величинами. Поэтому они определяются той наибольшей частью молекул, энергия которых близка к среднему значению энергии. Для чего же нам нужно знать, например, что есть ничтожная доля (порядка 0,00001 %) молекул, кинетическая энергия которых больше 17 kT? Ведь эти отдельные молекулы с очень большой энергией на величину давления и на общий запас энергии газа практически не влияют.

Картина, однако, меняется, если рассматривать *химические реакции*. Оказывается, что именно эти редкие молекулы с большой энергией полностью определяют ход химических реакций. Загадка химических реакций заключается в том, что молекулы, вступающие в реакцию, сталкиваются между собой каждые 10^{-10} с, а между тем реакция часто требует нескольких минут (а иногда и часов) для своего завершения. Значит, только редчайшая часть всех столкновений приводит к химической реакции.

Высказывались предположения, что у молекул есть какое-то очень маленькое «чувствительное место», в которое нужно попасть, чтобы произошла реакция. Эта точка зрения похожа на легенду о греческом герое Ахиллесе, единственным уязвимым местом которого была пятка.

Однако правильное объяснение, которое дал в конце XIX в. шведский ученый Сванте Аррениус, заключается в том, что к реакции приводят только столкновения молекул, энергия которых превышает определенное значение, так называемую *энергию активации* E_a.

Когда, например, сталкиваются молекулы водорода и иода и образуются две молекулы иодистого водорода, то нужно, чтобы энергия сталкивающихся молекул была больше $3\cdot10^{-19}$ Дж. Для сравнения отметим, что при 0° С величина $kT=1{,}38\cdot10^{-23}\cdot273\approx3{,}8\times10^{-21}$ Дж. Таким образом, при комнатной температуре (или при температуре 0° С, имеющей тот же порядок величины) нужной энергией обладает лишь ничтожная часть молекул $\alpha=e^{-\nu}$, где $\nu=3\cdot10^{-19}/3{,}8\cdot10^{-21}\approx80$, откуда получим $\alpha\approx1/10^{35}$.

Время релаксации получим, умножая время между двумя столкновениями (величина порядка 10^{-10} с) на то среднее число столкновений, среди которых встретится одно столкновение с нужной энергией. Это среднее число столкновений имеет порядок величины $\frac{1}{\alpha}$, т. е. близко к 10^{35}. Получим время реакции при 0° С порядка 10^{25} с $\approx3\cdot10^{17}$ лет. Этот результат согласуется с тем фактом, что при 0° С реакция $H_2+I_2=2HI$ практически ненаблюдаема.

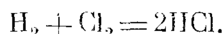
Из приведенных соображений следует, что в зависимости от температуры время реакции выражается формулой

$$t=\tau e^{E_a/kT},$$

где τ — время между двумя столкновениями, E_a — энергия активации. Эта формула правильно описывает зависимость скорости химических реакций от температуры. Характерная особенность формулы — крайне резкое уменьшение времени реакции и увеличение скорости реакции при небольшом изменении температуры.

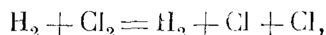
Часто, однако, в химических реакциях дело обстоит гораздо сложнее, потому что реакция может иметь различные промежуточные стадии.

В качестве примера рассмотрим реакцию

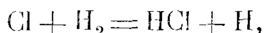


$$H_2+Cl_2=2HCl.$$

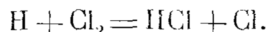
Эта реакция идет не путем столкновений молекул водорода с молекулой хлора, а по схеме



$$H_2+Cl_2=H_2+Cl+Cl,$$



$$Cl+H_2=HCl+H,$$



$$H+Cl_2=HCl+Cl.$$

В результате для фактически наблюдаемой скорости реакции получаются сложные зависимости. Однако для каждой отдельной реакции, например для

$$Cl+H_2=HCl+H \quad (\text{в реакции } H_2+Cl_2=2HCl),$$

имеет место закон Аррениуса — скорость реакции пропорциональна $e^{-E_a/kT}$, причем величина энергии активации E_a имеет различное значение для каждой реакции.

Советскому ученому академику Н. Н. Семенову принадлежит заслуга всестороннего исследования сложных (цепных) химических реакций, выяснения законов протекания таких реакций и общих причин, приводящих к протеканию реакции по сложной схеме.

§ 7. Испарение. Ток эмиссии катода

Идея Сванте Аррениуса о роли относительно малого числа молекул, энергия которых намного превосходит среднее значение энергии, помогает разобраться не только в химических реакциях, но и

в ряде других явлений. К таким явлениям относится *испарение жидкости*.

Для испарения жидкости нужно затратить значительную энергию. Так, например, для испарения 1 г воды при 100° С нужно затратить[1] около 2260 Дж. На одну молекулу это составляет[2] $Q = \frac{18 \cdot 2260}{6 \cdot 10^{23}} \approx 7 \cdot 10^{-20}$ Дж. Но при $T = 0°\,C = 273$ K $kT \approx 3{,}8 \times 10^{-21}$, поэтому $\frac{Q}{kT} \approx 20$. Оторваться от поверхности жидкости и испариться могут лишь те молекулы, энергия которых превышает теплоту испарения Q; доля таких молекул равна $e^{-Q/kT}$. Поэтому и скорость испарения пропорциональна $e^{-Q/kT}$. Заметим, что для удобства расчетов принято умножать числитель и знаменатель выражения $\frac{Q}{kT}$ на число Авогадро A:

$$\frac{Q}{kT} = \frac{QA}{kAT} = \frac{QA}{RT}.$$

Величина QA есть теплота испарения $6 \cdot 10^{23}$ молекул, т. е. теплота испарения 1 моля вещества. Величина $kA = R$ есть универсальная газовая постоянная: $R \approx 8{,}3$ Дж/(град·моль) (см. § 2). Теплота испарения 1 моля воды равна

$$Q_м \approx 18 \cdot 2260 \approx 40\,000 \text{ Дж/моль}.$$

Таким образом, скорость испарения воды пропорциональна

$$e^{-\frac{40\,000}{8{,}3T}} \approx e^{-\frac{5000}{T}}.$$

Рассмотрим насыщенный пар над поверхностью воды. Поскольку пар является насыщенным, число молекул воды, испаряющихся в единицу времени, равно числу молекул, находящихся в паре и прилипающих к поверхности воды (конденсирующихся) в единицу времени. Скорость испарения равна $Ce^{-Q_м/RT}$, где C — постоянная величина, пропорциональная площади поверхности воды. Скорость конденсации пропорциональна давлению паров воды и также пропорциональна площади поверхности воды. Следовательно, в случае насыщенного пара, когда скорости испарения и конденсации равны:

$$Dp = Ce^{-Q_м/RT},$$

где D и C — величины, пропорциональные площади поверхности, слабо зависящие от температуры и совсем не зависящие от давления, откуда

$$p = Fe^{-Q_м/RT},$$

где постоянная F не зависит от площади поверхности воды. Таким образом устанавливается связь между давлением насыщенного пара и теплотой испарения.

Рассмотрим еще один процесс, похожий на испарение, — *испускание электронов* нагретой поверхностью. Этот процесс осуществляется на катоде радиоламп. Холодный катод в пустоте не испускает электронов[3]. Однако при высокой температуре катод испускает электроны. При этом если на анод подан достаточно высокий положительный потенциал, то анод притягивает электроны, и каждый электрон, оторвавшийся от поверхности катода, попадает на анод. Электрический ток, протекающий в цепи через радиолампу, равен произведению числа электронов, испускаемых катодом в единицу времени, на величину заряда электрона.

Опыт показывает, что в этих условиях имеет место следующая зависимость силы тока j от температуры:

$$j = ge^{-Q/kT}.$$

Величина Q для различных катодов различна. Например, для катода из чистого вольфрама $\frac{Q}{k} = 55\,000°$, для катода из окиси бария $\frac{Q}{k} = 30\,000°$, и, следовательно, такой катод может работать при более низкой температуре. По зависимости j от T можно определить отношение $\frac{Q}{k}$. При этом величина Q, которая входит в последнюю формулу, совпадает с энергией, необ-

[1] Теплота испарения мало зависит от температуры: для воды $Q = 2260$ Дж/г при 100° С и 2500 Дж/г при 0° С. В дальнейшем мы этой зависимостью пренебрегаем.

[2] Относительная молекулярная масса воды равна 18; число Авогадро равно $6 \cdot 10^{23}$.

[3] Мы здесь не рассматриваем случай очень сильного электрического поля (10^6 В/см и больше), способного вырвать электроны даже из холодного катода. Не рассматривается также выбивание электронов из катода действием света или бомбардировкой катода электронами, ионами или другими частицами.

ходимой для того, чтобы вырвать электрон из катода [4].

Радиолампа дает замечательный способ измерения распределения по скоростям электронов, вылетающих с поверхности катода при данной температуре. Для этого при нагретом катоде дадим на анод небольшой отрицательный потенциал φ. При таком потенциале анод отталкивает электроны, испускаемые катодом. Поэтому большая часть электронов, не долетев до анода, повернет назад и упадет обратно на катод. Однако будут и такие электроны, которые все-таки попадут на анод, преодолев его отталкивание. Для этого нужно, чтобы кинетическая энергия электрона, вылетевшего с катода, превосходила разность потенциальной энергии у анода и у катода, т. е. величину $\varepsilon\varphi$, где под ε понимается заряд электрона (ведь буква e у нас, к сожалению, занята!). Доля этих электронов равна $e^{-\varepsilon\varphi/kT}$. Таким образом, при отрицательном потенциале анода φ ток равен $j=j_0e^{-\varepsilon\varphi/kT}$, где j_0 есть ток при положительном потенциале. При осуществлении опыта необходимо, чтобы расстояние между катодом и анодом было мало, для того чтобы количество электронов между ними не было велико и взаимное отталкивание электронов не повлияло на результат опыта.

Советский ученый академик А. Ф. Иоффе предложил использовать это явление для прямого превращения тепловой энергии в электрическую. Если электроны идут с катода на отрицательно заряженный анод, то в целом такая система является источником напряжения: во внешней цепи между отрицательно заряженным анодом и положительным катодом ток имеет такое направление, что он совершает работу. Такой способ получения электрического тока замечателен отсутствием каких бы то ни было движущихся частей, принципиальной простотой схемы: в этом отношении он похож на получение электроэнергии при помощи термоэлементов, также осуществленное А. Ф. Иоффе. В настоящее время полупроводники, в которых свободные электроны имеются уже при комнатной температуре, успешно заменили радиолампы в приемниках и вычислительных машинах. Однако накаленный катод, испускающий электроны по описанному выше закону, до сих пор остается существенной частью телевизионной трубки.

[4] Энергия, необходимая для вырывания электрона, может быть определена другими способами.

Глава 12

ПОГЛОЩЕНИЕ И ИЗЛУЧЕНИЕ СВЕТА. ЛАЗЕРЫ

§ 1. Поглощение света (постановка задачи и грубая оценка)

Рассмотрим поглощение света в воздухе, содержащем черные пылинки сажи. Пусть в единице объема содержится N пылинок. Площадь сечения одной пылинки плоскостью, перпендикулярной к лучу света, обозначим через σ; кратко σ будем называть просто *сечением*. Так, например, для пылинки, представляющей собой шар радиуса r, сечение σ есть площадь фигуры, по которой пересекает шар плоскость, проходящая через его центр, т. е. $\sigma=\pi r^2$ [1].

Будем считать, что свет, попадающий на поверхность пылинки, полностью поглощается. Задача заключается в определении зависимости доли поглощенного света и доли прошедшего света от величин N, σ и длины пути x, которую луч света прошел через воздух, содержащий пыль.

Начнем с самой грубой оценки того пути, на котором происходит поглощение значительной доли света. Эту длину обозначим через L. Что именно означает многозначительное выражение «значительная доля света» мы разберем в следующих параграфах. Не будем смущаться нечеткой постановкой вопроса, достаточно обычной в физических задачах, где сплошь и рядом уточнение задачи происходит позднее, в процессе ее решения.

Рассмотрим цилиндр с площадью основания S и образующей длины L. Потребуем, чтобы *сумма сечений всех заключенных в цилиндре пылинок равнялась* S.

[1] Для пылинок сложной формы точное определение сечения σ таково: σ есть с р е д н я я площадь тени, отбрасываемой пылинкой на поверхность, перпендикулярную к лучу света. При этом среднюю величину площади определяют, производя измерения площади тени при различных ориентациях пылинок относительно направления луча света, причем плоскость, на которую падает тень пылинки, при всех измерениях считается перпендикулярной лучу.

Каков физический смысл поставленного условия? Если бы можно было разместить пылинки так, чтобы закрытые разными пылинками площади не перекрывались, то при помощи пылинок, находящихся в цилиндре длины L с площадью основания S, можно было бы закрыть в с е основание цилиндра и достичь полного поглощения всего света. При $x<L$ полное поглощение света будет заведомо невозможно: как ни располагай пылинки, все равно их окажется недостаточно для того, чтобы обеспечить полное поглощение света.

Ясно, что на самом деле при $x=L$ и даже при $x>L$ полного поглощения не будет: при случайном расположении пылинок при л ю б о м x могут оставаться такие направления, вдоль которых не окажется ни одной пылинки, и лучи света свободно пройдут по этим направлениям.

В объеме $V=SL$ цилиндра содержится NSL пылинок, сумма их сечений равна σNSL; значит, мы требуем, чтобы было

$$\sigma NSL=S,$$

откуда

$$L=\frac{1}{\sigma N}. \qquad (12.1.1)$$

Проверим размерность в (1): σ есть площадь, ее размерность $см^2$; N есть число пылинок в единице объема, размерность $\frac{1}{см^3}$. Следовательно, $[L]=\frac{1}{см^2\cdot см^{-3}}=см$, как и должно быть.

Энергия, которую свет переносит через какую-нибудь площадку за 1 с, называется *потоком энергии* света. Пусть I есть поток энергии света, переносимой через площадку в 1 $м^2$. Эта величина называется *плотностью потока энергии*, ее размерность $Дж/м^2\cdot с$. Ниже мы будем рассматривать плотность потока световой энергии $I(x)$ в зависимости от толщины слоя x. Ясно, что

$$I(x)=I_0 f(x), \qquad (12.1.2)$$

где I_0 есть энергия падающего света, $f(x)$ — искомая функция, характеризующая ослабление света.

Что можно сказать о свойствах функции $f(x)$? Если $x=0$, то ослабления света еще не произошло, $I(0)=I_0$; поэтому $f(0)=1$. Если $x>0$, то происходит ослабление света, т. е. $I(x)<I_0$; поэтому $f(x)<1$. При этом ясно, что с увеличением x функция $f(x)$ уменьшается, стремясь к нулю, т. е. $f(x)$ — убывающая функция. Значит, ее производная отрицательна: $\frac{df}{dx}<0$.

Мы говорили, что полного поглощения не будет ни при $x=L$, ни при $x>L$; поэтому мы не ждем, чтобы $f(x)$ обратилась в нуль при $x=L$. Однако можно предполагать, что величина $x=L$ является характерной длиной. Это значит, что при прохождении светом пути $x \ll L$ доля поглощенного света весьма мала по сравнению с долей прошедшего света; при прохождении пути $x \approx L$ поглощается заметная часть света, а при прохождении пути $x \gg L$ бо́льшая часть света поглощается, проходит только весьма малая часть света.

Функция $f(x)$, как видно из формулы (2), безразмерна. Можно надеяться, что если ввести безразмерную переменную $\frac{x}{L}$, то функция $f\left(\frac{x}{L}\right)$ будет всегда одинакова для любых пылинок, любых N и σ. Эти предположения будут подтверждены и уточнены в следующих параграфах.

§ 2. Уравнение поглощения и его решение

Все расчеты мы по-прежнему ведем для столба воздуха в виде цилиндра с площадью основания 1 м^2 (в предыдущем параграфе рассматривался цилиндр с площадью основания S м^2, но в формуле (1.1) величина S все равно сократилась). Рассмотрим тонкий слой воздуха, отвечающий участку цилиндра между значениями x и $x+dx$ его длины.

Пучок света состоит из параллельных лучей и характеризуется плотностью потока энергии I. Если бы не было поглощения света пылинками, величина I была бы постоянной.

В рассматриваемом слое содержится Ndx пылинок, которые, если их тени не перекрываются, что при малом dx вполне правдоподобно, покрывают часть σNdx полной площади 1 м^2 основания слоя. Следовательно, в слое поглощается доля σNdx энергии, падающей на слой: $dQ=IN\sigma dx$. При прохождении светом слоя dx поток энергии света уменьшается на величину, равную количеству dQ поглощенной энергии. До входа в слой плотность потока энергии была $I(x)$, а после выхода из слоя она стала $I(x+dx)$; поэтому

$$I(x)-I(x+dx)=I\sigma Ndx. \qquad (12.2.1)$$

Но $I(x+dx)-I(x)\approx dI$; поэтому из (1) следует

$$\frac{dI}{dx}=-IN\sigma. \qquad (12.2.2)$$

Решение этого дифференциального уравнения, как мы знаем, имеет вид

$$I=I_0e^{-\sigma Nx} \qquad (12.2.3)$$

(ср. с формулами (6.6.10) и (6.6.9); см. также гл. 8). Здесь $I_0=I(0)$ есть значение плотности потока энергии при $x=0$.

При увеличении толщины слоя в арифметической прогрессии $x_1=a$, $x_2=2a$, $x_3=3a$ и т. д. сила света уменьшается в геометрической прогрессии. В самом деле, обозначая $e^{-\sigma Na}=\alpha$ (где, конечно, $\alpha<1$), найдем, пользуясь (3):

$$I(x_1)=I_0\alpha, \quad I(x_2)=I_0\alpha^2,$$

$$I(x_3)=I_0\alpha^3 \text{ и т. д.}$$

§ 3. Соотношение между точным и грубым расчетами поглощения

Весьма поучительно теперь, располагая и точным решением (§ 2) и грубой оценкой (§ 1), сравнить их. Такое сравнение помогает пользоваться грубыми оценками в сложных задачах, где точное решение найти трудно: сравнение помогает понять границы применимости грубого решения (ср. с § 10.6).

При грубом решении мы нашли длину, на которой происходит значительное поглощение: $L=\frac{1}{N\sigma}$.

При помощи «характерной длины» L точное решение (2.3) можно выразить так:

$$I=I_0e^{-\frac{x}{L}}. \qquad (12.3.1)$$

Значит, полностью подтверждается предположение, что найденная грубым

рассуждением величина L входит в точное решение.

Действительно, точное решение имеет вид

$$I = I_0 f\left(\frac{x}{L}\right),$$

где (ср. (1)) $f\left(\frac{x}{L}\right) = e^{-x/L}$.

Рассмотрим длину $x=L$. Грубое рассуждение давало на этой длине полное поглощение света. На самом деле из точного решения (1), полагая в нем $x=L$, находим $I=I_0e^{-1}\approx 0{,}37I_0$, т. е. при длине L цилиндра проходит 37% света, а поглощается, следовательно, 63%. При малых $\frac{x}{L}$ выразим $e^{-x/L}$ по приближенной формуле $e^{-u}\approx 1-u$ (см. § 4.8), сохраняющей лишь первые два члена разложения e^{-u} в ряд (ср. § 6.1); мы получим

$$e^{-x/L}\approx 1-\frac{x}{L}. \qquad (12.3.2)$$

Геометрически это соответствует замене кривой $y=e^{-x/L}$ на касательную к этой кривой в точке $x=0$ (рис. 1). Как видно из (2), касательная пересекает ось x при $x=L$. Поэтому если бы поглощение происходило с одинаковой скоростью, т. е. так, что на каждой единице длины поглащалось бы одинаковое количество света, а именно столько же, сколько в начале пути, то весь свет был бы поглощен на длине $x=L$.

Таким образом, величина L, полученная грубым рассуждением, действительно имеет важнейшее значение и в точном решении.

Подчеркнем еще раз, что умение находить (и применять) неточные, «прикидочные» решения очень важно в практической работе любого физика или инженера, — и надо всячески развивать умение быстро (так сказать, «прикидкой») находить приближенное решение задачи, тренироваться в нахождении и понимании таких решений: это гораздо важнее и плодотворнее выпячивания недостатков грубых решений. Будем радоваться тому, что совсем простая прикидка дает 100% поглощения там, где точное решение указывает, что поглощение равно 63%, т. е. что наша прикидка дает ошибку всего в 1,5 раза. (Конечно, такой же приближенный подход при $x=L$ дает 0% прохождения света вместо точного значения 37%, т. е. уменьшает прохождение света в бесконечное число раз, но и это не беда, так как здесь заранее было понятно, что от грубого решения нельзя ждать хорошей точности и «0% прохождения света» надо воспринимать лишь как указание на значительное его поглощение.)

Если установлено, что задача не имеет точного решения в виде явной формулы, то на этом никак нельзя останавливаться. Надо стараться отыскать хотя бы очень грубое решение задачи (не забывая, конечно, при пользовании им, что это не точное, а только приближенное, «прикидочное» решение).

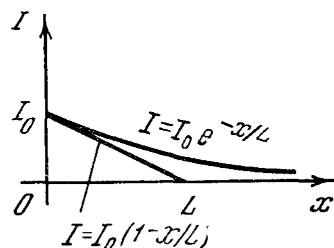


Рис. 12.3.1

Остановимся еще на вопросе о размерностях. Размерность величины $L=\frac{1}{N\sigma}$ мы проверили и установили, что это длина. Часто оказывается возможным найти приближенное выражение интересующей нас величины, зная только ее размерность и размерность исходных величин, входящих в условие задачи[1]. Однако в данном случае это невозможно. Действительно, величину, имеющую размерность длины, можно построить исходя из одной только концентрации $N\left(\frac{1}{\text{м}^3}\right)$: $l_1=\frac{1}{\sqrt[3]{N}}$. (Величина l_1 представляет собой *среднее расстояние* между частицами.) Величину, имеющую размерность длины, можно построить также исходя из сечения σ (м²): $l_2=\sqrt{\sigma}$ (величина l_2 характеризует *средний размер* частицы). Очевидно, что величина $l_\alpha=l_1^\alpha l_2^{1-\alpha}$ при любом значении показателя степени α также имеет размерность длины. В частности, интересующая нас величина l получается при $\alpha=3$.

[1] В этом заключается так называемый *метод размерностей* (или *теория размерности*); такой подход имеет широкие применения в физике, механике и технике.

Таким образом, в рассмотренной задаче соображения размерности не приводят к определенному ответу. Для того чтобы найти L, т. е. величину размерности длины, входящую в точное решение, оказывается необходимым найти именно грубое решение задачи. Но даже в тех случаях, когда соображения размерности дают единственный ответ на вопрос, желательно также найти грубое решение задачи, чтобы яснее представить картину явления.

Отметим, наконец, что расчеты § 2 основывались на предположении о случайном распределении пылинок при заданной средней их концентрации. Мы предполагаем, что на входе (при $x=0$) поток энергии q_0 строго одинаков по всему сечению. Однако после того как свет прошел определенную длину x, в воздухе, содержащем пылинки, строго говоря, появились участки сечения (общей площадью S_1), на которые упала тень от пылинок (и там $q=0$); остались также незатененные участки общей площадью S_2, где $q=q_0$ (при этом $S_1+S_2=S$, ибо общая площадь постоянна). Величина, которую мы вычисляем, есть поток, усредненный по всей поверхности:

$$q(x)=\frac{q_0 S_2}{S_1+S_2}=\frac{q_0 S_2}{S}.$$

При расчете изменения q по длине существенно предположение, что пылинки, находящиеся в слое между x и $x+dx$, распределены в этом слое *случайно*, т. е. не предпочитают специально освещенные или затемненные участки.

Продумайте сами вопрос о том, как изменится функция $q(x)$, если пылинки размещены в воздухе стройными рядами, как атомы в кристалле.

§ 4. Эффективное сечение

В задаче об ослаблении света, проходящего через воздух с пылью, величина σ имеет простой геометрический смысл площади тени одной пылинки. Закон ослабления света (2.2) одинаков для света различной длины волны (т. е. разного цвета), так как величина σ от длины волны не зависит.

При поглощении света отдельными молекулами и атомами, напротив, наблюдается сильная зависимость закона ослабления света от длины волны света. Так, например, в чистом воздухе при атмосферном давлении видимый свет практически совершенно не ослабляется (ослабление составляет меньше 1% на пути 1 км, соответственно ослабление в e раз происходит на пути около 100 км). Ультрафиолетовые лучи с длиной волны $1{,}8\cdot10^{-7}$м$=1800$ Å[1] ослабляются в e раз на пути $L=0{,}001$ м$=0{,}1$ см. Еще более короткие ультрафиолетовые лучи, с длиной волны $1{,}1\cdot10^{-5}$ см$=1100$ Å, ослабляются в e раз на пути $L=0{,}01$ см.

Следовательно, поглощение света воздухом не похоже на поглощение света черной пылинкой, одинаково поглощающей свет с любой длиной волны.

Количество световой энергии q, поглощаемое одним атомом в единицу времени, пропорционально плотности потока световой энергии I в том месте, где находится атом:

$$q=\sigma I. \qquad (12.4.1)$$

Здесь σ есть коэффициент пропорциональности. Определим размерность σ. Размерность q есть Дж/с. Размерность потока энергии I есть Дж/(м^2·с). Следовательно, размерность σ есть м^2. Величина σ называется *эффективным сечением*. Для черной пылинки коэффициент σ совпадает с геометрической величиной — площадью тени. Для молекул и атомов σ сильно зависит от длины волны света.

Грубо можно так представить себе причину этой зависимости. Количество поглощаемой атомом энергии при воздействии на него света оказывается особенно большим, когда частота световых колебаний совпадает с частотой движения электронов в атоме. При этом имеет место явление *резонанса*, электрон колеблется особенно сильно и поглощает особенно много световой энергии.

Такой резонанс достигается, например, при поглощении атомами натрия (в парообразном состоянии) желтого света с длиной волны 5890 Å$=5{,}89\cdot10^{-7}$ м.

Точно такой же желтый свет испускают атомы натрия при более высокой температуре, когда колебание

[1] Å — ангстрем, 1 Å$=10^{-10}$ м.

электронов вызывается сильными столкновениями атомов между собой.

При резонансе величина σ достигает 10^{-14} м², т. е. имеет порядок квадрата длины световой волны. Размеры атомов и молекул порядка 10^{-10}—10^{-9} м, что соответствует сечениям порядка 10^{-20}—10^{-18} м². Таким образом, максимальные эффективные сечения во много раз больше истинных площадей сечений атомов и молекул. С другой стороны, для света, частота которого не соответствует собственной частоте атома, эффективное сечение мало, гораздо меньше площади сечения атома.

Отметим, наконец, что в очень разреженном газе, как правило, атом, возбужденный светом, «переизлучает» свет и притом в произвольном направлении. В плотном газе возбужденный атом отдает свою энергию другим атомам при столкновении.

Таким образом, в разреженном газе имеет место рассеяние света, а в плотном газе — истинное поглощение. Однако для расчета уменьшения интенсивности *направленного* светового луча дальнейшая судьба отнятой энергии безразлична.

§ 5. Ослабление потока заряженных частиц α- и β-лучей

Экспоненциальный закон убывания потока частиц в зависимости от пройденного пути

$$I = I_0 e^{-x/L} \tag{12.5.1}$$

основан на очень общем предположении о том, что ослабление потока на малом пути dx пропорционально самой интенсивности потока:

$$\frac{dI}{dx} = -\frac{1}{L} I, \tag{12.5.2}$$

где коэффициент пропорциональности $\frac{1}{L}$ есть постоянное число, зависящее только от вида частиц. Так как (1) есть решение дифференциального уравнения (2), то очевидно, что формулы (1) и (2) равносильны.

Опыты показывают, что в некоторых случаях экспоненциальный (показательный) закон (1) является вполне точным, а иногда от него наблюдаются те или иные отклонения. Рассмотрим внимательно те причины, которые могут вызвать отклонения от формулы (1) или (что то же самое) от закона (2).

Легко ответить на вопрос о смысле отклонений от формулы (2). Соотношение (2) подразумевает, что при изменении x и I рассматриваемый свет (или какое-либо другое излучение) качественно не меняется, иначе изменялось бы число L. Перепишем (2) в виде

$$\frac{1}{I}\frac{dI}{dx} = -\frac{1}{L}.$$

Отсюда видно, что величина $\frac{1}{I}\frac{dI}{dx}$ постоянна. Если окажется, что в различных точках пространства величина $\frac{1}{I}\frac{dI}{dx}$ различна, значит, в этих точках различна не только интенсивность излучения, но и его физическое содержание: например, мы имеем свет разного цвета, т. е. с неодинаковой средней длиной волны.

При рассмотрении вопросов защиты от радиоактивных излучений и вопросов прохождения α, β, γ-лучей и нейтронов через различные вещества мы встречаемся с другой причиной отклонений от простого закона (2).

В применении к процессу поглощения света закон (2) означает следующее: если свет встретил на своем пути пылинку, то часть его проходит мимо этой пылинки, совсем не изменяясь, а другая часть света целиком поглощается пылинкой. В случае радиоактивных излучений дело обстоит сложнее: α-частица представляет собой ядро атома гелия, вылетающее из радиоактивного «родительского» ядра с большой скоростью порядка 0,07 c (c — скорость света), т. е. со скоростью около $2 \cdot 10^7$ м/с. Пролетая через атом, α-частица отдает электронам малую часть своей энергии. Примерно после 50 000 столкновений с атомами α-частица потеряет половину своей энергии. При этом она не перестанет существовать, не исчезнет, но энергия и скорость ее изменятся. После 100 000 столкновений α-частица останавливается, перестает сталкиваться с атомами и выбивать из них электроны. Остановившаяся частица отбирает от окружающих атомов два электрона и превращается в нейтральный покоящийся (за исключением общего теплового движения) атом гелия. Когда Резерфорд установил, что попадание α-частиц в сосуд с газом сопро-

вождается накоплением гелия в этом сосуде, то этим завершился важнейший этап исследования радиоактивности.

Число столкновений, необходимое для остановки, α-частица испытывает в воздухе на пути в несколько сантиметров. В действительности на этом пути различные α-частицы (с одинаковой начальной энергией), разумеется, испытывают разное число столкновений, не обязательно точно 100 000. Однако, так как 100 000 — это очень большое число, то при данной длине пути отклонения числа столкновений от среднего значения этого числа (примерно равного как раз 100 000) весьма невелики — порядка 300, т. е. составляют всего лишь около 0,3% от общего числа столкновений[1]. Поэтому α-частицы одинаковой энергии всегда теряют свою энергию на приблизительно одинаковом пути, который зависит от начальной энергии α-частицы. Если поток α-частиц одинаковой энергии летит вдоль оси x, то зависимость между интенсивностью потока и величиной пройденного пути x изображается кривой рис. 1. Кривая эта совершенно не похожа на график показательной функции. На протяжении значительной части пути интенсивность потока α-частиц не меняется: за одинаковые промежутки времени через площадку в 1 см^2 пролетает одинаковое число α-частиц. Затем интенсивность резко падает. Это резкое падение было подготовлено на участке, где интенсивность оставалась постоянной, потому что на этом участке энергия α-частиц уменьшалась с увеличением пути x. Резкое падение потока

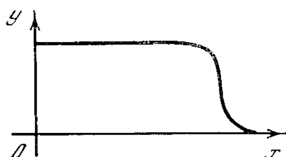


Рис. 12.5.1

[1] Здесь работает так называемый *закон больших чисел*, относящийся к многократному повторению однотипных *случайных событий*, исход каждого из которых непредсказуем. Этот закон утверждает, что при большом числе повторений доля случаев, когда будет иметь место определенный исход события, весьма мало отличается от ожидаемого (среднего) числа таких исходов (эта доля неизбежно стремится к среднему при неограниченном возрастании числа N повторений события).

происходит там, где энергия α-частиц становится весьма малой.

Аналогичная картина имеет место и для быстрых электронов (β-частиц, испускаемых, когда в ядре нейтрон превращается в протон). В этом случае картина осложняется тем, что при радиоактивном распаде испускаются электроны с различными скоростями и к тому же электроны, пролетая около атома, не только отдают ему часть своей энергии, но часто испытывают и значительное боковое отклонение.

Кривая зависимости интенсивности потока β-частиц от расстояния имеет вид, показанный на рис. 2, — она существенно отлична от кривой на рис. 1. Уже при небольших x часть электронов выбывает из пучка; это в основном те электроны, которые имели малую начальную скорость. Поэтому вблизи $x = 0$ поведение кривой похоже на поведение показательной функции. Однако дальше кривая достигает оси абсцисс, интенсивность становится равной нулю при вполне определенной величине x, соответствующей максимальной энергии электронов, получающихся при данном радиоактивном распаде.

Наиболее важными для практики (и вместе с тем наиболее трудными) являются вопросы, связанные с защитой от γ-лучей, испускаемых радиоактивными веществами, и от нейтронов, образующихся при делении ядер в атомных реакторах (котлах) и при ядерных взрывах. В этом случае положение наиболее запутанное и сложное, так как γ-лучи

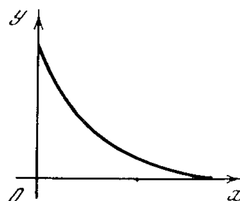


Рис. 12.5.2

и нейтроны отдают энергию большими порциями и при этом сильно отклоняются от первоначального направления. Даже в толстом слое воздуха (100—200 м) есть значительная вероятность (порядка 37%) прохождения неизмененных γ-лучей и нейтронов. Поэтому они требуют толстых защитных слоев. Поток γ-лучей и нейтронов не обращается в нуль при определенной толщине слоя, как это имеет место для α- и β-лучей: при большой толщине защитного слоя,

как показывают опыт и сложные расчеты, поток γ-лучей и нейтронов убывает приблизительно по экспоненциальному закону, причем показатель степени здесь часто отличается от того, который соответствует начальному участку кривой.

§ 6*. Поглощение и испускание света горячим газом

Представим себе горячий газ, через который проходит пучок света. При этом мы рассмотрим более подробно, что происходит с атомами, например атомами натрия, при поглощении и испускании света и при столкновениях с другими атомами и молекулами, учитывая высокую температуру газа.

Прежде всего следует отметить, что уподобление атома шарику на пружине, т. е. «осциллятору» — системе с определенной частотой собственных колебаний, во многих случаях является недостаточно точным. Атом — это система, состоящая из ядра и электронов и подчиняющаяся законам квантовой механики. Этих законов мы здесь не можем коснуться; для нас существенно лишь то, что такая система может находиться только в определенных состояниях и может иметь только строго определенные, раздельные (или, как выражаются математики, *дискретные*) значения энергии. Назовем нижнее состояние, характеризующееся наименьшим возможным значением энергии, «нулевым», «основным», или «невозбужденным»; для обозначения относящихся к нему величин мы будем пользоваться индексом нуль: энергия атома в этом состоянии есть E_0, число атомов в нулевом состоянии в единице объема (плотность этих атомов) равно n_0 и т. д. Точно так же можно говорить о первом (после нижнего) состоянии с энергией атома E_1, плотностью атомов n_1, о следующем — втором состоянии (энергия E_2, плотность n_2) и т. д. При этом мы считаем, что

$$E_0 < E_1 < E_2 < \ldots$$

При низкой температуре все атомы имеют наименьшую возможную энергию, т. е. находятся в низшем состоянии: их энергия равна E_0 (каждого!), а плотность n_0 не отличается от суммарной плотности n всех атомов (во всех возможных состояниях [1]).

Атом в низшем состоянии может поглощать свет, переходя при этом в другое состояние, например в первое возбужденное состояние. Прибегнем снова к квантовой теории. Можно сказать, что свет состоит из отдельных частиц, *фотонов*, и энергия каждого фотона равна $h\nu$, где h — *постоянная Планка*, равная $6{,}63\cdot10^{-34}$ Дж·с, а ν — *частота* света (число колебаний в 1 с). Особенность квантовой теории заключается в том, что, говоря о частицах света, мы не отказываемся и от таких понятий, как частота ν, длина волны и т. д., т. е. от понятий, связанных с представлением о световых волнах.

Но не будем останавливаться на философской стороне квантовой механики!

Из закона сохранения энергии следует, что если при поглощении света (одного фотона) атом переходит из нижнего в первое возбужденное состояние, то

$$E_0 + h\nu = E_1. \tag{12.6.1}$$

Это же равенство имеет место и при испускании света возбужденным атомом, переходящим из первого состояния в нулевое. Значит, резонансная частота ν поглощенного или испускаемого света зависит от разности энергий двух состояний атома: эта частота (которую можно определить с помощью спектроскопа) характеризует переход из одного состояния в другое. Величина ν — точнее было бы обозначить ее через $\nu_{1,0}$ — не есть частота колебаний электронов в атоме в каком-то одном (нулевом или первом) состоянии: она зависит от обоих состояний. Экспериментаторы, занимавшиеся спектроскопией, заметили эту закономерность давно, ибо она легко усматривается из таблицы частот многих отдельных линий спектра определенного атома. В самом деле, имеют место (почему? докажите это сами) равенства типа

$$\nu_{4,1} + \nu_{1,0} = \nu_{4,3} + \nu_{3,0}$$

и т. д., которые можно проверить, ничего не зная об атомах.

[1] Говоря более строго, $n = n_0 + n_1 + n_2 + \ldots$ Но при низкой температуре $n_1 \ll n_0$, $n_2 \ll \ll n_0$, ... и сумма $n_1 + n_2 + \ldots \ll n_0$, так что $n \approx \approx n_0$.

Нильс Бор первым понял истинный смысл и причину возникновения этих равенств. Он сформулировал свои «квантовые постулаты»:

1) существует набор состояний (0, 1, 2,. . .), находясь в которых, атом света не излучает;

2) излучение происходит при п е р е х о д е атома из одного состояния в другое, причем частота излучаемого света равна разности энергий исходного и конечного состояний, деленной на постоянную Планка.

Аналогичное положение характерно и для поглощения света.

Вернемся теперь к основной теме параграфа — к вопросу о поглощении и излучении света г о р я ч и м газом. Напомним, что при н и з к о й температуре газа все его атомы находятся в низшем состоянии, и свет резонансной частоты (мы теперь снова пишем кратко ν вместо $\nu_{1,0}$) поглощается, так что

$$\frac{dI}{dx} = -\sigma n_0 I, \tag{12.6.2}$$

где I — световой поток, а σ — так называемое сечение (см. (2.2)). Что же изменится в случае г о р я ч е г о газа?

При высокой температуре в результате теплового движения атомов и их столкновений между собой или с атомами и молекулами других газов (если мы имеем дело со смесью газов) часть атомов перейдет в состояние с более высокой энергией. Другими словами, теперь будут иметься и возбужденные атомы, т. е. величинами n_1, n_2, . . . уже нельзя будет пренебрегать. Согласно общему закону теплового движения (см. § 11.4) при данной температуре T

$$n_0 : n_1 : n_2 : \ldots = e^{-E_0/kT} : e^{-E_1/kT} : \\ : e^{-E_2/kT} : \ldots \tag{12.6.3}$$

Если задано общее число n атомов в единице объема, то из (3) и равенства $n = n_0 + n_1 + n_2 + \cdots$ следует

$$n_0 = n \frac{e^{-E_0/kT}}{e^{-E_0/kT} + e^{-E_1/kT} + \ldots} = \\ = \frac{n}{1 + e^{-(E_1-E_0)/kT} + \ldots},$$

$$n_1 = \frac{n e^{-(E_1-E_0)/kT}}{1 + e^{-(E_1-E_0)/kT} + \ldots}, \quad \ldots \tag{12.6.4}$$

Значит, с повышением температуры при данном n поглощение уменьшается, ибо $n_0 < n$. Но, кроме того, — и это самое главное — появляются возбужденные атомы, способные излучать свет. При этом они излучают свет той же самой частоты, что и поглощаемый свет.

Можно предположить, что уравнение для интенсивности света при наличии возбужденных атомов будет иметь вид

$$?\frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1.? \tag{12.6.5}$$

Здесь коэффициент $\varkappa$ должен иметь смысл произведения вероятности w того, что возбужденный атом в единицу времени перейдет в невозбужденное (нулевое) состояние, испуская фотон света, на вероятность α того, что испущенный фотон присоединится к рассматриваемому потоку светового излучения. Поскольку поток I выражается в энергетических единицах Дж/(м$^2\cdot$с), то надо еще умножить $w\alpha$ на энергию $h\nu$ одного фотона. Можно рассуждать так: в слое dx сечением в 1 см^2 находится $n_1 dx$ возбужденных атомов — и ограничиться проверкой размерностей: w имеет размерность 1/с$=$с$^{-1}$, α безразмерно; n_1 имеет размерность м$^{-3}$, так что если $\varkappa = w\alpha h\nu$, то $\varkappa n_1$ имеет размерность Дж/(м$^3\cdot$с), т. е. такую же размерность, как и $\frac{dI}{dx}$. Контроль сравнением размерностей как будто убеждает, что формула (5) правильна. И все же эта формула неверна!

Вопросительные знаки в начале и конце (5) как раз и поставлены для того, чтобы рассеянный читатель не счел бы эту формулу правильной. В том, что ее нужно изменить, мы убедимся далее, рассматривая *термодинамическое равновесие* световых потоков.

§ 7*. Термодинамическое равновесие излучения

Предположим, что сосуд, в котором находится газ, имеет идеально отражающие зеркальные стенки. Температура газа задана и поддерживается постоянной.

Естественно предположить, что если в сосуде первоначально не было света, то в результате излучения возбужденными атомами он затем появится. Если в сосуде в начальный момент времени

интенсивность света слишком велика, то лишний свет будет поглощен атомами, находящимися в нижнем невозбужденном состоянии.

Можно ожидать, что каким бы ни было начальное состояние, в результате испускания и поглощения света газом заданной температуры установится вполне определенная интенсивность света. Более тонкие рассуждения показывают, что интенсивность света данной частоты зависит только от температуры. В самом деле, представим себе два разных сосуда, наполненных разными газами, имеющими одну и ту же температуру. Соединим наши сосуды световодом со спектральным фильтром, пропускающим туда и обратно только свет одной определенной частоты. Если бы интенсивности света были различны, то, несмотря на одинаковую температуру сосудов, один сосуд отдавал бы энергии больше, чем получал, и охлаждался бы, а другой сосуд нагревался бы, что явно невозможно.

Итак, существует определенная *равновесная интенсивность* света — та, при которой устанавливается равновесие между испусканием и поглощением. Эта равновесная интенсивность при данной частоте зависит только от температуры, а не от свойств газа. Естественно поэтому пытаться найти равновесную интенсивность излучения исходя непосредственно из общих принципов теплового движения.

Рассмотрим излучение определенной частоты и определенного направления в сосуде. Энергия этого излучения не может принимать любые значения: согласно квантовой теории (т. е. в данном случае — согласно теории фотонов) энергия может равняться нулю (ни одного фотона, отвечающего данной длине волны), или равняться $h\nu$ (один фотон), или равняться $2h\nu$ (два фотона) и т. д.

По общим принципам отношения вероятностей $p_0, p_1, p_2, \ldots$ соответствующих состояний таковы:

$$p_0 : p_1 : p_2 : \ldots = 1 : e^{-h\nu/kT} : e^{-2h\nu/kT} : \ldots \quad (12.7.1)$$

в соответствии с тем, что энергии E здесь равны 0, $h\nu$, $2h\nu$, ..., а вероятности пропорциональны $e^{-E/kT}$; $e^0 = 1$, поэтому последовательность в правой части (1) и начинается с единицы. Но сумма всех вероятностей равна единице, потому что мы исчерпали все возможности — ведь данный луч должен характеризоваться либо одним, либо двумя и т. д. фотонами, либо отсутствием фотонов. Значит, мы можем, как говорят, «нормировать» распределение (1), добавить к этим равенствам, задающим отношения вероятностей, еще равенство

$$p_0 + p_1 + p_2 + \ldots = 1, \quad (12.7.2)$$

указывающее, чему равна сумма всех вероятностей. Равенства (1), очевидно, дают:

$$p_n = ae^{-nh\nu/kT}; \quad n = 0,\ 1,\ 2,\ \ldots \quad (12.7.3)$$

(т. е. $p_0 = a$, $p_1 = ae^{-h\nu/kT}$ и т. д.), где множитель a пока остается неопределенным. Найти его можно из (2):

$$p_0 + p_1 + p_2 + \ldots =$$
$$= a(1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + \ldots) =$$
$$= \frac{a}{1 - e^{-h\nu/kT}}, \quad (12.7.4)$$

и, значит,

$$a = 1 - e^{-h\nu/kT} \quad (12.7.4а)$$

(заметьте, что в (4) в скобках стоит сумма членов геометрической прогрессии).

Теперь мы уже можем найти *среднее число* фотонов φ и *среднюю энергию* (точнее, *плотность энергии*) ε данного сорта лучей:

$$\varphi = 1 \cdot p_1 + 2 \cdot p_2 + \ldots, \quad (12.7.5)$$

$$\varepsilon = \varphi h\nu \quad (12.7.5а)$$

(так как φ — это среднее число фотонов в единичном объеме, то размерность этого числа — $м^{-3}$; размерность ε — Дж/м^3). Подставляя (3) и (4а) в (5) (а затем (5) в (5а)), получаем:

$$\varphi = \frac{e^{-h\nu/kT}}{1 - e^{-h\nu/kT}} = \frac{1}{e^{h\nu/kT} - 1}, \quad (12.7.6)$$

$$\varepsilon = \frac{h\nu}{e^{h\nu/kT} - 1} \quad (12.7.6а)$$

(здесь опущены несложные выкладки и сразу приведен окончательный результат [1]). Данной частоте ν соответствует

[1] Для того чтобы вывести из (3), (4а) и (5) результат (6), достаточно заметить, что $\varphi - e^{-h\nu/kT}\varphi = p_1 + p_2 + p_3 + \ldots = 1 - p_0$ $(= 1 - a;$ почему?).

определенный период $\frac{1}{\nu}$ колебания и определенная длина $\lambda = \frac{c}{\nu}$ световой волны.

Можно найти число различных волн в объеме V, частота которых заключена между ν и $\nu + d\nu$. Нам надо учесть, что электромагнитные волны имеют определенную поляризацию: электрическое поле перпендикулярно направлению распространения волны (но в трехмерном пространстве можно указать много взаимно перпендикулярных направлений, перпендикулярных данному!). Учитываются и условия отражения света от стенок, но оказывается, что число независимых видов колебаний, частота которых заключена в данном интервале, в первом приближении зависит только от объема V и пропорционально этому объему. Таким образом, можно говорить о числе колебаний dK в интервале частот $d\nu$, приходящихся на единицу объема; это число оказывается равно

$$dK = \frac{8\pi\nu^2 d\nu}{c^3}, \tag{12.7.7}$$

где c есть скорость света. При этом так как частота ν (а значит, и $d\nu$) имеет размерность $с^{-1}$, а c — размерность м/с, то размерность dK есть $м^{-3}$.

Полный вывод формулы (7) слишком громоздок для того, чтобы его можно было включить в пособие для начинающих. Однако приближенно понять результат (7) и убедиться в его разумности нетрудно. Заметим, что $\frac{\nu}{c} = \frac{1}{\lambda}$, откуда $\nu = \frac{c}{\lambda}$ и $|d\nu| = \frac{c}{\lambda^2} d\lambda$. Значит, по (7) число колебаний с длиной волны, заключенной между λ и $\lambda + d\lambda$, равно

$$dK = \frac{8\pi}{\lambda^4} d\lambda. \tag{12.7.7a}$$

Число же колебаний с длиной волны, заключенной, скажем, между λ_0 и $2\lambda_0$, равно

$$K_{\lambda_0, 2\lambda_0} = 8\pi \int\limits_{\lambda_0}^{2\lambda_0} \frac{d\lambda}{\lambda^4} = \frac{8\pi}{3}\left(\frac{1}{\lambda_0^3} - \frac{1}{2^3\lambda_0^3}\right) = \frac{7\pi}{3}\,\frac{1}{\lambda_0^3}. \tag{12.7.8}$$

Таким образом, если отвлечься от численного множителя, по порядку величины не сильно отличающегося от единицы, то можно сказать, что на единицу объема приходится столько же колебаний с длиной волны, близкой к λ_0, сколько можно уместить в этом объеме кубиков с ребром λ_0. Каждое колебание в среднем «занимает объем порядка λ_0^3» — результат, не только легко запоминающийся, но и вполне естественный.

Теперь можно легко написать выражение Q полной *энергии излучения* в единице объема. Для этого возьмем энергию каждого колебания с определенной частотой, умножим на число колебаний данного сорта и сложим (т. е. проинтегрируем) все эти энергии:

$$Q = \int \varepsilon\, dK = \int\limits_0^\infty \frac{h\nu}{e^{h\nu/kT} - 1}\,\frac{8\pi\nu^2 d\nu}{c^3} \tag{12.7.9}$$

(Q имеет размерность плотности энергии — Дж/м^3). Удобно обозначить $\frac{h\nu}{kT} = x$ и выделить безразмерный интеграл:

$$Q = \frac{8\pi h}{c^3}\left(\frac{kT}{h}\right)^4 \int\limits_0^\infty \frac{x^3 dx}{e^x - 1}. \tag{12.7.9a}$$

Фигурирующий в (9а) неопределенный интеграл не выражается через элементарные функции. Приближенно интеграл

$$I = \int\limits_0^\infty \frac{x^3 dx}{e^x - 1} \tag{12.7.10}$$

можно вычислить, замечая, что

$$\frac{1}{e^x - 1} = e^{-x}\frac{1}{1 - e^{-x}} = e^{-x}(1 + e^{-x} + e^{-2x} + \ldots) = e^{-x} + e^{-2x} + \ldots,$$

и, значит,

$$\int \frac{x^3 dx}{e^x - 1} = \int x^3 e^{-x} dx + \int x^3 e^{-2x} dx + \ldots \tag{12.7.11}$$

Но (см. (7.5.19))

$$\int\limits_0^\infty x^3 e^{-x} dx = 3\,! = 6,$$

$$\int\limits_0^\infty x^3 e^{-2x} dx = \frac{1}{2^4}\int\limits_0^\infty y^3 e^{-y} dy = \frac{1}{16}\cdot 6 \text{ и т. д.},$$

так что

$$\int_0^\infty \frac{x^3 dx}{e^x - 1} = 6\left(1 + \frac{1}{16} + \frac{1}{81} + \ldots\right). \tag{12.7.11a}$$

Стоящий в правой части (11a) ряд сходится; ограничиваясь пятью членами этого ряда, мы получим для интеграла (10) приближенное значение 6,482.

С помощью теории функций комплексной переменной (см. § 17.2) можно получить точное значение определенного интеграла (10). Оно равно $\frac{\pi^4}{15}$, т. е. примерно 6,494; таким образом, допущенная при ограничении пятью членами ряда (11) ошибка меньше 0,2%.

Итак, точное выражение равновесной плотности излучения (в Дж/м^3) при данной абсолютной температуре T (по шкале Кельвина) есть

$$Q = \frac{8\pi^5}{15} \frac{k^4 T^4}{c^3 h^3} = 7{,}5 \cdot 10^{-16} T^4. \tag{12.7.12}$$

Представим себе, что сосуд, в котором установилась такая плотность излучения, имеет плоское отверстие в одной из стенок площадью S квадратных сантиметров. Тогда через это отверстие будет вытекать энергия. Поток энергии будет равен RS, где R измеряется в Дж/(м^2·с) и равно

$$R = \frac{c}{4} Q = 5{,}75 \cdot 10^{-8} T^4. \tag{12.7.13}$$

Электромагнитная энергия «течет», распространяется со скоростью света; но в каждый данный момент в единице объема внутри сосуда только половина фотонов движется в сторону отверстия, а другая половина движется в противоположную сторону. Да и те фотоны, которые движутся в сторону отверстия, летят под разными углами — и это еще уменьшает поток вдвое. При точном учете этих обстоятельств в формуле (13) и возникает коэффициент 1/4.

Заткнем отверстие в нашем ящике с зеркальными стенками каким-нибудь черным материалом, притом нагретым до такой же температуры, которую имеют газ и излучение в ящике.

Когда мы говорим, что материал *черный* (или «абсолютно черный»), то это значит, что поглощается любой фотон, с любой частотой, упавший на поверхность из нашего материала. Наша пробка должна была бы нагреваться, поглощая излучение. Но если пробка нагрета сама до той же температуры, что и излучение, то она, очевидно, не будет нагреваться дальше, до более высокой температуры. Значит, пробка будет излучать ровно столько же, сколько поглощает. Таким образом, мы приходим к выводу, что черная поверхность с температурой T излучает с единицы поверхности $R = 5{,}67 \cdot 10^{-8}\, T^4$ Дж/(м^2·с). Эта величина выражает то, что называют «излучением абсолютно черного тела».

Избранный нами порядок изложения является обратным по отношению к тому, как исторически развивалось исследование излучения. Сперва был экспериментально найден закон излучения (12), т. е. установлен тот факт, что излучение пропорционально четвертой степени температуры, причем коэффициент пропорциональности равен $5{,}67 \cdot 10^{-8}$ Дж·м$^{-2}$·с$^{-1}$ · град$^{-4}$ (з а к о н С т е ф а н а — Б о л ь ц м а н а, 1888 г.).

Затем было найдено спектральное распределение излучения — ф о р м у л а П л а н к а (6а) (1899 г.). В этой формуле впервые появилась величина $h = 6{,}63 \cdot 10^{-34}$ Дж·с — так называемая *постоянная Планка*.

Для света с данной частотой характерна определенная энергия $h\nu$. Планк считал, что поглощение и излучение происходят определенными порциями, а к распространению света «порционная», или «квантовая», теория отношения не имеет. Позже (1905 г.) Эйнштейн показал, что свет состоит из частиц с определенной энергией, равной $h\nu$, и импульсом, равным $\frac{h\nu}{c}$. Наконец, в 1924 г. индийский ученый Бозе дал тот вывод формулы Планка, исходящий из расчета вероятностей $p_0, p_1, \ldots$, с которого мы и начали изложение.

§ 8*. Вероятность излучения и условия термодинамического равновесия

Вернемся к выражению для интенсивности луча света, пропускаемого горячим газом.

Из уравнения (6.5) (напомним, что это уравнение неправильно!)

$$?\ \frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1\ ?$$

и из условия $\frac{dI}{dx}=0$ получим *равновесную* интенсивность $I_{рв}$:

$$?\, I_{рв}=\frac{\varkappa n_1}{\sigma n_0}\,.? \qquad (12.8.1)$$

Отношение концентраций n_1 возбужденных атомов и n_0 атомов в основном состоянии равно $\frac{n_1}{n_0}=e^{-\frac{E_1-E_0}{kT}}$ (см. (6.3) или (6.4)). Согласно квантовому условию (6.1) $E_1-E_0=h\nu$, значит, $\frac{n_1}{n_0}=$ $=e^{-h\nu/kT}$, так что

$$?\, I_{рв}=\frac{\varkappa}{\sigma}\,e^{-h\nu/kT}.\ ? \qquad (12.8.1а)$$

Формула (1а) для $I_{рв}$ похожа на правильную формулу (7.6а), которая была получена Планком, но не совпадает с ней.

Отметим сперва именно черты сходства: равновесная плотность излучения $I_{рв}$ не зависит от концентрации газа, атомы которого поглощают (в основном состоянии) и излучают (в возбужденном), а зависит лишь от частоты ν света и от температуры T. Вид зависимости в пределе, т. е. при отношении $\frac{h\nu}{kT}\gg 1$, совпадает с тем, который был у Планка. В самом деле, если $\frac{h\nu}{kT}\gg 1$, то $e^{h\nu/kT}\gg 1$ — и в формуле Планка (7.6а)

$$\varepsilon=\frac{h\nu}{e^{h\nu/kT}-1}$$

можно пренебречь единицей в знаменателе по сравнению с (большим!) слагаемым $e^{h\nu/kT}$: при $h\nu\gg kT$

$$\varepsilon\approx\frac{h\nu}{e^{h\nu/kT}}=h\nu e^{-h\nu/kT},$$

откуда вытекает, что

$$I_{рв}\approx ch\nu e^{-h\nu/kT},\quad \frac{h\nu}{kT}\gg 1, \qquad (12.8.2)$$

ведь I — это энергия в единицу времени, а c — скорость света. (Заметим, что поскольку плотность энергии ε имеет размерность Дж/м^3, а c — размерность м/с, то произведение $c\varepsilon$ имеет размерность Дж/(м$^2\cdot$с) потока энергии I.)

Итак, полученная из наивных предположений формула (1а) по характеру функциональной зависимости $I_{рв}$ от частоты ν и температуры T не отличается от точной формулы (2). Для того чтобы совпадение было полным, нужно, чтобы отдача энергии возбужденным атомом в одиночную волну, характеризуемая коэффициентом $\varkappa$, находилась с σ-сечением поглощения света атомом в нормальном состоянии в соотношении, определяемом равенством

$$\frac{\varkappa}{\sigma}=ch\nu. \qquad (12.8.3)$$

С точки зрения учения о резонансе (да и с точки зрения квантовой механики) такое соотношение представляется вполне естественным. Если мало́ сечение поглощения света (например, потому, что мал заряд колеблющегося шарика), то мала и вероятность излучения, также зависящая от заряда шарика.

Перейдем от вероятности излучения единичной волны к полной вероятности излучения энергии возбужденным атомом. А именно, разделив количество энергии, излучаемой в единицу времени в одну волну, на энергию $h\nu$ фотона, мы получим вероятность излучения в одну волну. Затем полученные величины складываем, т. е. интегрируем по всем значениям длины (или частоты) волны [1]:

$$w=\int\frac{\varkappa}{h\nu}\,\frac{8\pi\nu^2 d\nu}{c^3}=\int\frac{\sigma 8\pi\nu^2 d\nu}{c^2} \qquad (12.8.4)$$

(размерность w есть 1/с=с$^{-1}$). Здесь w есть вероятность излучения, так что для концентрации возбужденных атомов в отсутствие внешнего излучения имеет место равенство

$$\frac{dn_1}{dt}=-wn_1. \qquad (12.8.5)$$

Решение уравнения (5) очевидно:

$$n_1(t)=n_1(t_0)\,e^{-w(t-t_0)}. \qquad (12.8.6)$$

Соответственно $\tau=w^{-1}$ есть среднее время жизни возбужденного атома (ср. § 8.2).

Мы рассматривали лишь п е р в о е возбужденное состояние, которое может распадаться только одним способом, переходя в основное (невозбужденное) состояние. Как правило, при

[1] Не бойтесь того, что в интеграл формально входят частоты, которые данный возбужденный атом не может испускать. Для этих частот $\varkappa=0$; множитель $\varkappa(\nu)$ в подынтегральной функции отсечет их.

данном основном состоянии и данном строго определенном возбужденном состоянии (с определенной пространственной ориентацией атома) сечение поглощения σ зависит от направления и поляризации падающей волны. В этом случае точная формула имеет вид

$$w = \int (\sigma_1 + \sigma_2) \frac{\nu^2 d\nu}{c^2} d\Omega, \qquad (12.8.7)$$

где σ_1 и σ_2 относятся к двум поляризациям волны, $d\Omega$ есть элемент телесного угла. При $\sigma_1 = \sigma_2 = \sigma$, не зависящих от направления, получим $\sigma_1 + \sigma_2 = 2\sigma$ и $\int d\Omega = 4\pi$, что и даст множитель 8π в упрощенной формуле (4).

Обратите внимание на размерность: σ имеет размерность м^2, а $\frac{c}{\nu} = \lambda$ — длина волны — размерность м, так что в интеграле

$$w = \int \frac{8\pi\sigma}{\lambda^2} d\nu \qquad (12.8.8)$$

величина $\frac{\sigma}{\lambda^2}$ безразмерна. Существенно то, что σ сильно зависит от частоты: поглощение света является резонансным процессом. Вспомним теорию колебаний: поглощение энергии осциллятором при учете трения зависит от частоты по формуле

$$A \sim \frac{\text{const}}{(\nu - \nu_0)^2 + \gamma^2}. \qquad (12.8.9)$$

Оказывается, что в случае покоящегося атома поглощение света подчиняется аналогичной формуле [2]:

$$\sigma = \frac{\lambda^2}{2\pi} \frac{\gamma^2}{(\nu - \nu_0)^2 + \gamma^2}. \qquad (12.8.9а)$$

Максимальное сечение поглощения света имеет величину порядка квадрата длины световой волны! Для видимого света эта величина во много раз больше размеров самого атома: желтая линия натрия имеет длину волны $\lambda = 5400 \times \times 10^{-10}$ м, а радиус орбиты электрона порядка 10^{-10} м, сечение резонансного поглощения фотона в миллионы раз больше площади, ограниченной орбитой электрона. Здесь проявляется тот факт, что фотон нельзя рассматривать как точечную частицу, которая летит по прямой и либо попадает в мишень, либо пролетает мимо, не задевая мишени. Согласно квантовой механике движение фотона определяется волновыми уравнениями электромагнитного поля, — и с этим связана возможность захвата фотона атомом с расстояния порядка длины световой волны.

До сих пор мы подчеркивали и углубляли сходство нашей приближенной формулы (1а) для равновесной плотности излучения с подтвержденной опытом точной формулой Планка (2). Теперь обратимся к отличию формулы Планка от приближенного выражения.

Как надо видоизменить выражение $\frac{dI}{dx}$ для того, чтобы равновесная плотность соответствовала формуле Планка? Ответ на этот вопрос дал Эйнштейн в 1917 г. Как это часто бывает в физике, дать ответ оказалось легче, чем правильно поставить вопрос.

Ответ Эйнштейна заключался в следующем. Заменим уравнение (6.5)

$$\frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1$$

следующим:

$$\frac{dI}{dx} = -\sigma n_0 I + \varkappa n_1 (1 + \varphi), \qquad (12.8.10)$$

где согласно предыдущему $\varkappa = ch\nu\sigma$. Если φ есть среднее число фотонов в данном луче (см. § 7), то

$$I = ch\nu\varphi, \qquad \frac{dI}{dx} = ch\nu \frac{d\varphi}{dx}. \qquad (12.8.11)$$

Подставляя (11) в (10), получим

$$\frac{dI}{dx} = ch\nu \frac{d\varphi}{dx} =$$

$$= ch\nu\sigma [-n_0\varphi + n_1 (1 + \varphi)]. \qquad (12.8.12)$$

Прежде всего убедимся, что поставленная цель достигнута. В самом деле, условие $I = I_{\text{рв}}$, т. е. $\frac{dI}{dx} = 0$, приводит к следующему результату для равновесного количества фотонов:

$$\frac{\varphi_{\text{рв}}}{1 + \varphi_{\text{рв}}} = \frac{n_1}{n_0} = e^{-h\nu/kT},$$

т. е. $-n_0\varphi_{\text{рв}} + n_1 (1 + \varphi_{\text{рв}}) = 0$,

или

$$\varphi_{\text{рв}} = \frac{e^{-h\nu/kT}}{1 - e^{-h\nu/kT}} = \frac{1}{e^{h\nu/kT} - 1}; \qquad (12.8.13)$$

тем самым мы пришли к выражению для вероятности поглощения и излучения, приводящему к формуле Планка.

---

[2] Подставляя (9а) в интеграл, стоящий в правой части (8), получим $w = 4\pi\gamma$.

Теперь оглянемся и подумаем, какой ценой получен этот результат.

Мы добавили к вероятности изучения $\varkappa n_1$ еще один член $\varkappa n_1\varphi$, равный $\sigma n_1 I$. Таким образом, мы (вслед за Эйнштейном) утверждаем, что существует процесс испускания фотонов, вызванный наличием пучка фотонов и пропорциональный числу фотонов, или интенсивности излучения уже существующего пучка. Процесс этот называют *индуцированным излучением*. Слово «индуцированное» переводится как «вызванное» — речь идет об излучении, вызванном наличием света.

Индуцированное излучение возникает наряду со *спонтанным*, т. е. самопроизвольным излучением; в выражении для излучения в скобках стоят два слагаемых $(1+\varphi)$, или, если раскрыты скобки, $\varkappa n_1 + \sigma n_1 I$, которые и отвечают этим двум видам излучения. Соответственно и уравнение для изменения числа возбужденных атомов имеет вид

$$\frac{dn_1}{dt} = -wn_1 - \frac{\sigma n_1 I}{h\nu} + \frac{\sigma n_0 I}{h\nu}. \qquad (12.8.14)$$

При отсутствии излучения у нас остается только член $-wn_1$, как мы и писали раньше. Но при наличии излучения происходит не только перескок атомов из основного состояния в возбужденное — так сказать, снизу вверх $\left(\text{этому отвечает член } \frac{\sigma n_0 I}{h\nu} \text{ в правой части (14)}\right)$, но и дополнительное индуцированное излучение с «переходом атомов вниз» $\left(\text{член} - \frac{\sigma n_1 I}{h\nu} \text{ в том же равенстве (14)}\right)$.

Подчеркнем особо, что сечение σ в членах, описывающих поглощение $\left(\frac{\sigma n_0 I}{h\nu}\right)$ и индуцированное излучение $\left(-\frac{\sigma n_1 I}{h\nu}\right)$, одинаково. Если интенсивность пучка I велика и членом n_1, описывающим спонтанное испускание, можно пренебречь по сравнению со слагаемыми, выражающими поглощение и индуцированное испускание, то уравнение для I упрощается:

$$\frac{dI}{dx} = -\sigma(n_0 - n_1) I. \qquad (12.8.15)$$

Очевидно, что условием для этого является «сильное» неравенство $I \gg I_{\text{рв}}$, где $I_{1\text{в}}$ — равновесная интенсивность излучения при температуре газа.

Подставим в (15) равновесную концентрацию возбужденных атомов $n_1 = n_0 e^{-h\nu/kT}$; тогда мы получим формулу для эффективного поглощения:

$$\frac{dI}{dx} = -\sigma(1 - e^{-h\nu/kT}) n_0 I. \qquad (12.8.16)$$

При нагреве газа эффективное поглощение, т. е. коэффициент при I, уменьшается не только вследствие уменьшения n_0, но еще и вследствие противодействующего влияния индуцированного излучения, что дает дополнительный множитель в скобках в правой части (16). Если $h\nu \ll kT$, то приближенно

$$e^{-h\nu/kT} \approx 1 - \frac{h\nu}{kT} \quad \text{и} \quad 1 - e^{-h\nu/kT} \approx \frac{h\nu}{kT} \ll 1;$$

уменьшение поглощения здесь очень значительно. Между тем, если есть только два уровня: 0 и 1, так что полное число атомов $n = n_0 + n_1$, то величина n_0 может уменьшиться не более чем в 2 раза при повышении T от 0 до ∞ [3].

Так, например, в радиоастрономии имеют дело с переходом атома водорода из основного состояния в слабовозбужденное с испусканием радиоволн с $\lambda = 21$ см. Энергия соответствующих фотонов равна $h\nu = \frac{hc}{\lambda} \approx 6{,}5 \cdot 10^{-34} \times 3 \cdot 10^8/0{,}21 \approx 10^{-24}$ Дж. Если температура водородного облака равна 100 К, то $kT \approx 1{,}37 \cdot 10^{-23} \cdot 100 \approx 10^{21}$ Дж — и множитель в скобках (в формуле (16)) таков: $(1 - e^{-h\nu/kT}) \approx \frac{h\nu}{kT} \approx 10^{-3}$ (ср. § 4.8).

Существенны качественные особенности индуцированного излучения. При этом процессе возбужденный атом излучает, усиливая проходящую (индуцирующую) волну, — иначе не могло бы компенсироваться поглощение этой волны атомами в нижнем (невозбужденном) состоянии. В этом отношении индуцированное излучение резко отличается от спонтанного, при котором возбужденный атом излучает во все стороны и только частично — в направлении проходящей волны.

---

[3] Правильнее здесь, конечно, было бы говорить о повышении температуры от 0 до $T \gg \frac{E_1 - E_0}{k}$, так как предельный переход $T \to \infty$ физически бессмыслен: атомы при этом распадутся.

На языке классической механики можно было бы сказать, что поглощение излучения подобно резонансной раскачке осциллятора (ср. § 10.5). Индуцированное излучение подобно торможению осциллятора: ведь и для того, чтобы затормозить колебания, нужна сила, находящаяся в резонансе с колебаниями, т. е. имеющая ту же частоту. Разница между силой раскачивающей и силой тормозящей осциллятор заключается в том, что отношения фаз колебаний сил к фазе осциллятора будут различными.

Однако никакая классическая аналогия не может полностью объяснить явление индуцированного излучения. Точная и полная теория индуцированного излучения возможна лишь на основании квантовой механики и квантовой теории электромагнитного излучения (квантовой теории поля). Теория индуцированного излучения органически входит в квантовую теорию излучения, основы которой заложены П. Дираком в 1927 г.

Тем большее восхищение вызывает интуиция Эйнштейна, который в 1917 г. сумел выявить все основные черты индуцированного излучения исходя из чисто термодинамических соображений, т. е. примерно так, как это было изложено выше. Напомним, что и само понятие о частицах света — фотонах — было впервые введено Эйнштейном в 1905 г.; именно за это он получил Нобелевскую премию.

§ 9*. Лазеры

Индуцированное излучение открывает замечательную возможность получения электромагнитных колебаний (в частности, света).

Представим себе среду, в которой концентрация n_1 возбужденных атомов больше концентрации n_0 атомов в основном состоянии. В обычных условиях теплового равновесия, при $\frac{n_1}{n_0} = e^{-h\nu/kT}$, такая ситуация невозможна, ибо здесь $\frac{n_1}{n_0} < 1$. Но существует несколько искусственных приемов, с помощью которых можно этого добиться.

Об этих приемах мы расскажем позже; пока же, не вдаваясь в подробности, примем, что $n_1 > n_0$, так что точному уравнению (8.10) можно придать вид

$$\frac{dI}{dx} = bI + \varkappa n_1, \qquad (12.9.1)$$

где $b = \sigma(n_1 - n_0) > 0$. Изменение знака коэффициента при I в выражении для $\frac{dI}{dx}$ самым радикальным образом влияет на характер решения. Соответствующее изменение свойств решения можно сравнить с изменением решения для концентрации нейтронов в процессе деления при переходе от подкритической массы делящегося вещества (урана 235, плутония) к надкритической массе (см. конец гл. 8).

В самом деле, если $b > 0$ в уравнении для $\frac{dI}{dx}$, то световой поток по мере распространения увеличивается экспоненциально. Пренебрегая в (1) вторым членом, найдем

$$I = I_0 e^{bx} > I_0, \qquad (12.9.2)$$

где I_0 — интенсивность света в начале координат, т. е. при $x=0$ (интенсивность света на входе в среду). Определенный объем газа (вещества) с $b > 0$, например трубка длины L, работает как *усилитель*: впуская поток I_0, получим на выходе поток $I_L = I_0 e^{bL} > I_0$, более сильный по сравнению с входящим потоком.

Подставим теперь с двух концов трубки зеркала с коэффициентами отражения β, δ. Если с левого конца ($x=0$) поступает поток I_0 и на правом конце выходит поток $I_L = I_0 e^{bL}$, то этот поток отразится в правом зеркале и на правый конец снаружи попадает поток

$$|I'_L| = \beta I_L = I_0 \beta e^{bL}, \qquad (12.9.3)$$

поэтому поток надо считать *отрицательным* (энергия переносится в сторону, противоположную направлению возрастания координаты). Для того чтобы не следить за знаком, будем отмечать такой поток штрихом сверху (будем писать I') и рассматривать его абсолютную величину.

Уравнение для I' имеет вид

$$\frac{d|I'|}{dx} = -bx. \qquad (12.9.4)$$

Это уравнение интегрируем «справа налево», задав в качестве «начального условия» величину $|I'_L| = I_0 \beta e^{bL}$, отвечающую значению $x=L$ (см. (3)), и на-

ходя отсюда величину I_0', отвечающую значению $x=0$:

$$|I_0'| = |I_L'| e^{-b\int\limits_L^0 dx} = |I_L'| e^{bL} = = I_0\beta e^{2bL}. \qquad (12.9.5)$$

После отражения от левого зеркала (коэффициент отражения δ) получается поток (рис. 1), равный

$$I_0'' = \beta\delta e^{2bL} I_0. \qquad (12.9.6)$$

Теперь учтем, что полный поток, идущий слева направо, равен

$$I_{\text{полн}} = I_0 + I_0'',$$

где I_0 — интенсивность падающего света, а I_0'' — дважды отраженного.

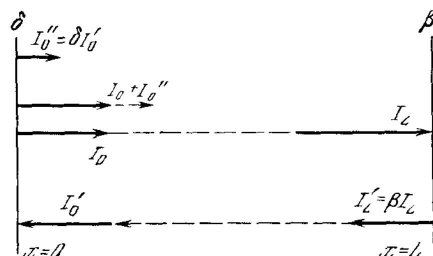


Рис. 12.9.1

При этом в правой части выражения (6) для I_0'' надо подставить именно $I_{\text{полн}}$, а не I_0 (мы не отличали I_0 падающего света от $I_{\text{полн}}$ выше, пока не учитывали двукратное отражение). При этом получим

$$I_{\text{полн}} = I_0 + I_0'' = I_0 + \beta\delta e^{2bL} I_{\text{полн}},$$

откуда

$$I_{\text{полн}} = \frac{I_0}{1 - \beta\delta e^{2bL}}. \qquad (12.9.7)$$

Физический вывод из всего сказанного заключается в том, что при $\beta\delta e^{2bL} = 1$ система, состоящая из трубки с активным газом и двух зеркал (с коэффициентами отражения β, δ), находится в «критическом» состоянии: здесь возможно стационарное состояние с любым значением I_0. Если величина $\beta\delta e^{2bL} < 1$, то система «подкритическая» (ср. с § 8.9). Без учета спонтанного излучения есть только одно решение $I_0 = 0$; с учетом спонтанного излучения (т. е. члена $\varkappa n_1$ уравнения (1)) в системе устанавливается излучение, пропорциональное $\varkappa n_1 L$. Наконец, при учете индуцированного излучения усиление будет более сильным; это усиление пропорционально отношению $\frac{1}{1 - \beta\delta e^{2bL}}$.

Такая ситуация технически мало интересна. Замечательная ситуация возникает, если, напротив, $\beta\delta e^{2bL} > 1$. По аналогии с цепной ядерной реакцией здесь можно говорить о «надкритическом» состоянии системы. В этом случае при любой малой начальной интенсивности и малом спонтанном излучении произойдет экспоненциальный рост интенсивности излучения в зависимости от времени, примерно пропорциональный $e^{\gamma t}$, где $\gamma \approx \frac{c}{2L} \frac{1}{\beta\delta e^{2bL} - 1}$. Интенсивность излучения будет расти, но одновременно будут расходоваться и возбужденные атомы; n_1 и b будут уменьшаться до тех пор, пока система не станет критической, но это произойдет лишь при большом значении интенсивности I.

Итак, в системе с $b > 0$ возможно получение большой интенсивности излучения. Это излучение строго монохроматично, представляет собой одно определенное колебание, поэтому такое излучение можно сфокусировать в точку. Полученное таким путем лазерное излучение обладает свойствами луча из пророческого научно-фантастического романа Алексея Толстого «Гиперболоид инженера Гарина»!

Применения лазерного луча описаны в огромном числе популярных и непопулярных статей, брошюр и книг. Здесь, в учебнике математики, мы лишь бегло укажем два способа создания инверсной заселенности, или активной среды — так физики называют ситуацию, когда $n_1 > n_0$, т. е. когда $b > 0$; эта ситуация позволяет получить излучение весьма большой интенсивности (приборы, основанные на использовании этого эффекта, называют *лазерами* или *мазерами*).

Один способ — исторически первый, которым воспользовались лауреаты Нобелевской премии: академики Н. Г. Басов и А. М. Прохоров и американский физик Ч. Таунс, — заключается в том, что молекулы, находящиеся в состояниях 1 и 0, поступают в конденсатор, помещенный в вакуум; здесь они движутся без столкновений, только под влиянием электрического поля. Электрические свойства, а именно поляри-

зуемость молекул в состояниях 0 и 1, различны; пользуясь этим, создатели лазера сумели обеспечить условия, при которых молекулы, находящиеся в основном состоянии, отклоняются и не попадают в сосуд, а молекулы в состоянии 1 в сосуд попадают. Таким путем в сосуде создается нужная нам ситуация, при которой $n_1 > n_0$.

Другой прием — более удобный и не требующий высокого вакуума, но исторически более поздний — заключается в том, что используется система с несколькими уровнями: 0, 1, 2, 3. Система «работает», т. е. создает излучение, связанное с переходом на один из промежуточных уровней, например с переходом $3 \to 2$, частоты $\nu_{3,2} = (E_3 - E_2)\, h$. Для этого возбуждают состояние 3, т. е. создают большую концентрацию n_3. При этом остается $n_3 < n_0$; возбудить молекулы более сильно (достичь того, чтобы было $n_3 > n_0$) путем облучения газа частотой $\nu_{3,0}$ нельзя из-за того же индуцированного излучения (продумайте сами, почему это так).

Однако если молекулы с уровня 2 очень быстро переходят на нижние уровни: $2 \to 1$ или $2 \to 0$, то вполне возможна ситуация, когда будет $n_2 \ll n_0$, $n_2 < n_3$, т. е. когда система будет надкритической по излучению частоты $\nu_{3,2}$.

Есть и другие способы получения активной среды. Общим для всех лазерных устройств является превращение энергии в форму монохроматической плоской электромагнитной волны. В этой форме энергия может быть сфокусирована, т. е. собрана в весьма малый объем. При высокой концентрации энергии можно достичь сверхвысоких температур и давлений, появляется надежда осуществить термоядерный синтез. Уже сейчас с помощью лазерного излучения удается осуществлять хирургические операции, технологические процессы, химические реакции — но все это, как уже было сказано, тема совсем других книг.

Глава 13

ЭЛЕКТРИЧЕСКИЕ ЦЕПИ И КОЛЕБАТЕЛЬНЫЕ ЯВЛЕНИЯ В НИХ

§ 1. Основные понятия и единицы измерения

В этой главе рассматриваются явления, происходящие в *электрических цепях*. Основными элементами электрической цепи являются *сопротивления*, *емкости*, *индуктивности*, *источники тока* (*напряжения*). Не претендуя на замену учебника физики, а лишь стремясь несколько дополнить, развить и уточнить знания, которые дает школьный учебник, напомним вкратце определения сопротивления, емкости и т. п. и единицы измерения этих величин; при этом мы считаем, что в объеме школьного курса физики все эти основные понятия читателю уже известны.

Количество электричества q определяется как разность количеств положительных и отрицательных зарядов. Единицей количества электричества является *кулон* (сокращенно Кл). Элементарный заряд — заряд протона — равен $e_p=1{,}6\cdot10^{-19}$ Кл; заряд электрона $e_e=-1{,}6\cdot10^{-19}$ Кл.

Сила тока (или, кратко, *ток*) определяется как количество электричества, протекающее в единицу времени через поперечное сечение проводника. Силу тока будем обозначать через j. За единицу силы тока принимается *ампер* (обозначение — А): ток в 1 А соответствует тому, что через поперечное сечение проводника за секунду протекает 1 Кл. Таким образом, 1 А=1 Кл/с; удобнее, впрочем, говорить, что 1 Кл=1 А·с, поскольку в системе СИ основной единицей является ампер, а кулон о п р е д е л я е т с я, как А·с.

За (положительное) *направление тока* принимается направление, в котором должны были бы двигаться п о л о ж и т е л ь н ы е заряды для того, чтобы создать данный ток. В действительности в металлических проводниках положительные заряды неподвижны, а ток течет благодаря движению электронов, несущих отрицательные заряды. Тело, заряженное положительно, — это, как правило, тело, которое потеряло часть своих электронов (только в редких случаях положительный заряд есть результат того, что тело приобрело положительные заряды). Отрицательно заряженное тело — это тело, которое приобрело избыточные электроны. Направление тока есть направление, п р о т и в о п о л о ж н о е тому направлению, в котором в проводнике движутся электроны.

Электрическим потенциалом данной точки называется потенциальная энергия, которой обладает положительный заряд в 1 Кл, помещенный в данную точку. При этом считается, что электрический потенциал земли равен н у л ю. Следовательно, равен нулю потенциал той точки схемы, которая соединена с землей металлическим проводом («заземлена»). Единицей потенциала служит *вольт* (обозначение — В). Потенциал точки равен 1 В, если заряд в 1 Кл, помещенный в эту точку, обладает потенциальной энергией в 1 Дж. (Напомним, что единица энергии или работы джоуль может быть определена как работа силы 1 Н на пути в 1 м: 1 Дж=1 Н·м=1 кг·м²/с².) Потенциальная энергия u заряда q, помещенного в точку, где потенциал равен φ, есть

$$u\,(\text{Дж})=q\,(\text{Кл})\cdot\varphi\,(\text{В}).\qquad(13.1.1)$$

При этом надо представить себе, что q мало, так как если поместить в данную точку большой заряд (например, 1 Кл), то изменится сам потенциал φ. Поэтому лучше сказать, что потенциал φ — это коэффициент при q в формуле (1) — коэффициент пропорциональности потенциальной энергии заряда и количества электричества.

Работа A, которую совершит поле, переводя заряд из точки, где потенциал равен φ_1, в точку, в которой потенциал равен φ_2, есть

$$A=u_1-u_2=q\,(\varphi_1-\varphi_2).$$

Подобно тому как во все физически значимые утверждения механики входит только р а з н о с т ь потенциальных энергий, но никогда — сама энергия, так и в учении об электричестве в формулы входит всегда р а з н о с т ь п о т е н ц и а л о в. Ко всем потенциалам во всех точках можно прибавить одинаковое слагаемое — разность потенциалов от этого не изменится. Поэтому потенциал в какой-нибудь одной точке схемы или аппарата всегда можно выбрать произвольно, например положить равным нулю, — и лишь после этого потенциалы всех других точек становятся вполне определенными. Именно это обстоятельство позволяет принять за нуль потенциал Земли.

Рассмотрим *конденсатор* (рис. 1), состоящий из двух параллельных пластин A и B. Одна из пластин (A) может быть

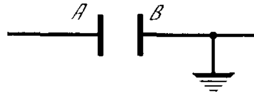


Рис. 13.1.1

соединена с каким-либо источником напряжения. Количество электричества на ней прямо пропорционально разности потенциалов φ_C пластин конденсатора:

$$q_A = C\varphi_C,$$

где разность потенциалов φ_C — это потенциал пластины A минус потенциал пластины B. Так как на рис. 1 пластина B заземлена, то здесь φ_C равно потенциалу пластины A.

Коэффициент пропорциональности C называется *емкостью* конденсатора. За единицу емкости принимается *фарада* (сокращено Ф) — емкость конденсатора с разностью потенциалов пластин 1 В при заряде 1 Кл (10^{-6} Ф=1 мкФ (микрофарада), 10^{-9} Ф=1 нФ (нанофарада), 10^{-12} Ф=1 пФ (пикофарада)).

На пластине B конденсатора скапливается равное по величине и противоположное по знаку количество электричества:

$$q_B = -q_A = -C\varphi_C.$$

Электрический заряд замкнутой системы есть сохраняющаяся величина: ни при каких процессах электрические заряды одного знака не возникают и не исчезают [1]. Изменение заряда пластины A конденсатора связано с тем, что заряд с пластины уходит куда-то в другое место, например в точку D (рис. 2).

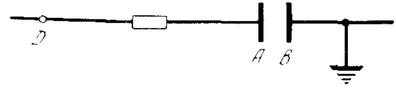


Рис. 13.1.2

Если в направлении от D к A (слева направо) течет ток j, то за время dt через поперечное сечение проводника пройдет количество электричества jdt; поэтому

$$dq_A = jdt, \quad \text{или} \quad \frac{dq_A}{dt} = j.$$

Рассмотрим теперь, от чего зависит сила тока, текущего по проводнику. По закону Ома сила тока пропорциональна разности потенциалов на концах проводника, причем ток течет от большего потенциала к меньшему. Таким образом,

$$j = k(\varphi_D - \varphi_A) = \frac{1}{R}(\varphi_D - \varphi_A). \qquad (13.1.2)$$

Положительная величина k называется *проводимостью*.

Обратная k величина $\frac{1}{k}$ называется *сопротивлением* проводника и обозначается буквой R. Единица сопротивления называется *ом* (обозначение — Ом) — это есть сопротивление проводника, по которому течет ток в 1 А при разности потенциалов на концах проводника в 1 В, так что 1 Ом=1 В/А.

Величину $\varphi_D - \varphi_A$ в формуле (2) уместно обозначить через φ_R — это есть разность потенциалов на сопротивлении R, причем величина φ_R (как выше φ_C) определяется как разность значений потенциала слева и справа (при этом ток течет слева направо). В этих обозначениях закон Ома (2) можно записать так:

$$j = \frac{\varphi_R}{R}, \quad \text{или} \quad \varphi_R = Rj. \qquad (13.1.3)$$

Источником напряжения в цепи может быть, например, гальванический

[1] При возникновении или исчезновении двух противоположно заряженных частиц с равным по абсолютной величине зарядом полный электрический заряд системы не меняется.

элемент. На полюсах элемента имеется определенная разность потенциалов. Приближенно можно считать, что эта разность потенциалов не зависит от тока, протекающего через элемент. В частности, в элементе ток может течь в сторону от более низкого потенциала к более высокому. Через сопротивление ток всегда течет от более высского потенциала к более низкому, подобно тому как по трубе, соединяющей два сосуда, вода течет из сосуда, в котором уровень воды выше, в сосуд, в котором уровень воды ниже.

Элемент подобен насосу, который может забирать воду в нижнем сосуде и поднимать ее в верхний сосуд, т. е. заставлять воду передвигаться снизу вверх. Для работы насоса необходим какой-то внешний источник энергии, приводящий насос в движение. Аналогичной является и ситуация в гальваническом элементе. Когда через элемент проходит ток от более низкого к более высокому потенциалу, в нем происходят химические реакции. Энергия этих химических реакций в элементе превращается в электрическую энергию.

Разность потенциалов, которую дает элемент, называется *электродвижущей силой*, или, сокращенно, ЭДС.

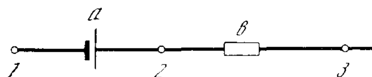


Рис. 13.1.3

Разность значений потенциалов на элементе слева и справа (рис. 3) равна ЭДС элемента со знаком минус:

$$\varphi_1 - \varphi_2 = -E.$$

В действительности ЭДС несколько зависит от силы тока, протекающего через элемент. Когда ток течет (на рис. 3 слева направо) в направлении от более низкого потенциала к более высокому (т. е. в нормальном режиме работы элемента, когда он производит электрическую энергию), ЭДС E уменьшается с увеличением силы тока. Приближенно можно считать ЭДС постоянной, но более точной является линейная зависимость E от силы тока:

$$E = a - bj, \qquad (13.1.4)$$

где (положительные) коэффициенты a и b являются характеристиками элемента. Элемент, ЭДС E которого н е з а в и с и т от силы тока j (т. е. элемент, для которого $b=0$ в формуле (4)), будем называть *идеальным*.

Рассмотрим последовательное соединение идеального элемента с ЭДС, равной a, и сопротивления b (см. рис. 3). Тогда

$$\varphi_E = \varphi_1 - \varphi_2 = -a, \quad \varphi_b = \varphi_2 - \varphi_3 = bj;$$

поэтому

$$\varphi_1 - \varphi_3 = (\varphi_1 - \varphi_2) + (\varphi_2 - \varphi_3) =$$
$$= -a + bj = -(a - bj) = -E.$$

Таким образом, мы вновь пришли к формуле (4) для ЭДС E. Поэтому величину b в формуле (4) часто называют *внутренним сопротивлением* элемента; реальный элемент, с которым мы обычно имеем дело, дает такую же зависимость (4) E от j, как соединенные последовательно идеальный элемент и сопротивление b. За величиной a сохраняют название ЭДС элемента и в случае реального (а не идеального) элемента (здесь имеется в виду, что $E=a$ при $j=0$ [2]; величина b характеризует падение ЭДС при $j \neq 0$).

В дальнейшем при рассмотрении электрических цепей, в которые входят источник тока, например гальванический элемент, и различные сопротивления, можно представить себе, что мы имеем дело с идеальным элементом с постоянной, не зависящей от тока ЭДС, а внутреннее сопротивление b источника объединить с внешним сопротивлением R. Таким образом, реальный элемент с внутренним сопротивлением b, включенный последовательно с сопротивлением R, эквивалентен идеальному элементу, включенному последовательно с сопротивлением $R_1 = R + b$. Ток в цепи в этом случае определяется формулой

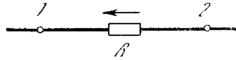
$$j = \frac{E}{R + b}.$$

Следует еще раз обратить особое внимание на различие между сопротивлением и источником напряжения. Если в какой-то цепи на сопротивлении имеется разность потенциалов такая, что $\varphi_2 > \varphi_1$ (рис. 4), то, по нашему определению, $\varphi_R = \varphi_1 - \varphi_2 < 0$, т. е. φ_R от-

[2] Так как при разомкнутой цепи ток равен нулю, то ЭДС можно определить как разность потенциалов разомкнутого элемента.

рицательно; следовательно, по формуле (3) и ток отрицателен. Значит, ток течет справа налево, от точки *2* к точке *1*. Теперь пусть разность потенциалов такого же знака имеется на концах источника напряжения, причем зависимость

Рис. 13.1.4



E от j дается формулой (4) (см. рис. 3). Пусть при этом $\varphi_3 > \varphi_1$, но $\varphi_3 - \varphi_1 < a$. Тогда $bj = \varphi_1 - \varphi_3 + a = a - (\varphi_3 - \varphi_1) > 0$, т. е. $j > 0$. Поэтому ток течет слева направо, несмотря на то что потенциал φ_1 слева меньше, чем потенциал φ_3 справа. Таким образом, источник напряжения способен преодолевать разность напряжений и давать положительный ток (слева направо) при отрицательной разности потенциалов ($\varphi_1 - \varphi_3 < 0$), если только эта отрицательная разность потенциалов по абсолютной величине не превышает ЭДС источника. Между тем сопротивление при отрицательной разности потенциалов всегда дает отрицательный ток. В любой схеме, состоящей только из сопротивлений (без ЭДС), мы будем иметь лишь тождественно равное нулю решение задачи об определении тока; при наличии же элементов (ЭДС; рис. 5) возможно и ненулевое решение.

Рассмотрим теперь явление *индуктивности*. Оно связано с магнитным полем, возникающим в окружающем проводник пространстве, когда по проводнику течет ток. Это магнитное поле особенно велико, если проводник имеет вид катушки с большим числом витков. Поле еще больше увеличивается, если катушка намотана на железный сердечник.

В свою очередь магнитное поле (переменное поле!) вызывает электрические явления. Известно, что каждый

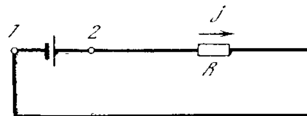


Рис. 13.1.5

виток (и даже каждая часть витка) катушки при наличии переменного магнитного поля становится источником напряжения, подобным гальваническому элементу. В катушке, в которой витки намотаны так, что ток обходит сердечник катушки в одном и том же направлении на всем протяжении катушки, все эти источники напряжения соединены последовательно, так что напряжения складываются.

В целом катушка эквивалентна источнику напряжения с разностью потенциалов, пропорциональной скорости изменения магнитного поля. Но магнитное поле в катушке пропорционально силе тока в катушке [3]. Поэтому скорость изменения магнитного поля пропорциональна с к о р о с т и изменения тока, т. е. п р о и з в о д н о й $\frac{dj}{dt}$. Окончательно в катушке (рис. 6)

$$\varphi_L = \varphi_1 - \varphi_2 = L\frac{dj}{dt}, \tag{13.1.5}$$

причем за положительное направление тока принято направление внутри катушки от точки *1* к точке *2*, а величина

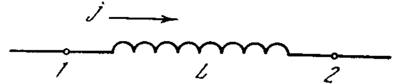


Рис. 13.1.6

φ_L есть разность потенциалов на катушке: потенциал φ_1 слева минус потенциал φ_2 справа. Рассматривая подробно направление магнитного поля и ЭДС, индуцируемую при его изменении, можно доказать, что коэффициент L в этой формуле (так называемая *индуктивность*) всегда положителен.

Из формулы (5) следует, что если $\frac{dj}{dt} < 0$, то $\varphi_1 - \varphi_2 < 0$, т. е. $\varphi_2 > \varphi_1$. Таким образом, если ток положителен (течет от *1* к *2*) и по величине уменьшается, то катушка играет роль элемента, поддерживающего в цепи положительный ток, несмотря на то что $\varphi_L < 0$. Если же ток положителен и увеличивается, то $\frac{dj}{dt} > 0$; поэтому $\varphi_L > 0$. В этом случае катушка играет роль добавочного сопротивления, так как здесь разность потенциалов на катушке положительна при положительном токе (ср. с (3)).

[3] Мы не будем рассматривать случай, когда на одном сердечнике намотаны две катушки, т. е. трансформатор, связывающий между собой две электрические цепи, в которых текут разные токи. Не рассматриваются также случаи более сложной зависимости магнитного поля от тока, когда в катушку вставлен железный сердечник и сила тока так велика, что железо «насыщается».

Существенное отличие катушки от источника напряжения и от сопротивления заключается в том, что величина φ_L зависит не от величины тока j, а от скорости $\frac{dj}{dt}$ его изменения.

Единица индуктивности (коэффициент L в уравнении (5); ранее вместо выражения «индуктивность катушки»[4] употреблялся термин «самоиндукция») называется *генри* (обозначение — Г). Утверждение, что индуктивность катушки

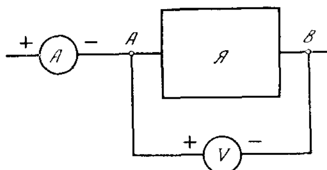


Рис. 13.1.7

равна 1 Г, означает, что при скорости изменения тока, равной 1 А/с, в катушке возникает разность потенциалов в 1 В. Размерность индуктивности получаем из формулы (5):

1 Г = 1 В · с/А.

Из сказанного видно, что индуктивность влияет на ток в цепи так же, как инертная масса (маховик) влияет на скорость движения: индуктивность препятствует изменению тока, масса (по второму закону Ньютона) препятствует изменению скорости. Подробнее эта аналогия будет рассмотрена в § 4.

С точки зрения дальнейших расчетов емкость, сопротивление, ЭДС и индуктивность имеют между собой нечто общее: в схеме все они присоединяются двумя проводами (в отличие, например, от трансформатора, имеющего четыре вывода, или радиолампы, имеющей три вывода — анод, катод, управляющая сетка). Приборы, которые включены в схему при помощи двух проводов, носят название *двухполюсников*; приборы, включаемые при помощи четырех проводов, называются *четырехполюсниками* и т. д. В схеме каждый элемент — емкость, сопротивление, ЭДС, индуктивность — в каждый момент времени характеризуется определенным током, который проходит через этот элемент, и определенной разностью потенциалов на входе и выходе.

Можно представить себе закрытый ящик *Я* (рис. 7) с двумя торчащими из него проводами *А* и *В*. Внутри ящика может быть что угодно: *R*, *E*, *L*, *C*. Подключим к ящику амперметр *А* и вольтметр *V*. При включении, показанном на рис. 7 (знаки «+» и «—» соответствуют надписям на клеммах амперметра и вольтметра), амперметр показывает ток j, идущий в направлении от *А* к *В*. Вольтметр показывает разность потенциалов

$$\varphi_я = \varphi_A - \varphi_B.$$

От того, что находится внутри ящика, зависит связь между $\varphi_я$ и j:

в случае сопротивления R $\quad \varphi_я = Rj,$ **(13.1.6)**

в случае ЭДС[5] E_0 $\quad \varphi_я = -E_0,$ **(13.1.7)**

в случае индуктивности L $\quad \varphi_я = L\frac{dj}{dt},$ **(13.1.8)**

в случае емкости[6] C $\quad \varphi_я = (\varphi_я)_0 + \frac{1}{C}\int_{t_0}^{t} j\,dt,$ **(13.1.9)**

или $\frac{d\varphi_я}{dt} = \frac{1}{C} j.$ **(13.1.9а)**

В некоторых случаях осуществляются и более сложные зависимости. Так, например, *выпрямитель тока* (двухэлектродная лампа или полупроводниковый

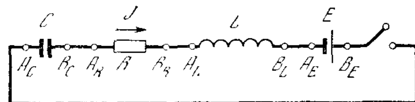


Рис. 13.1.8

диод) не подходит ни под одну из формул (6)—(9). Однако в большом числе важных вопросов можно ограничиться рассмотрением элементов схем, для ко-

[4] Часто вместо того, чтобы сказать «катушка, индуктивность которой равна *L*», для краткости говорят просто «индуктивность *L*». Говорят также «емкость *C*» вместо «конденсатор, емкость которого равна *C*». Точно так же говорят об ЭДС *E* вместо того, чтобы говорить о гальваническом элементе, или источнике напряжения.

[5] Внутреннее сопротивление ЭДС мы не учитываем.

[6] $q_A = C\varphi_я$, $\frac{dq_A}{dt} = j$, откуда $\frac{d\varphi_я}{dt} = \frac{1}{C} j$. Если в начальный момент времени $t = t_0$, $\varphi_я = (\varphi_я)_0$, то $\varphi_я = (\varphi_я)_0 + \frac{1}{C}\int_{t_0}^{t} j\,dt$.

торых с хорошей точностью выполняются формулы (6)—(9). Ниже мы будем рассматривать именно такие схемы, за исключением § 17, где специально рассматриваются свойства контура, в который включено устройство со сложной зависимостью тока от разности потенциалов.

Рассмотрим схему, изображенную на рис. 8. Выпишем падения напряжений на отдельных элементах схемы:

$$\varphi_C = \varphi_{A_C} - \varphi_{B_C}, \quad \varphi_R = \varphi_{A_R} - \varphi_{B_R},$$
$$\varphi_L = \varphi_{A_L} - \varphi_{B_L}, \quad \varphi_E = \varphi_{A_E} - \varphi_{B_E}. \quad (13.1.10)$$

Заметим еще, что $\varphi_{B_C} = \varphi_{A_R}$, $\varphi_{B_R} = \varphi_{A_L}$, $\varphi_{B_L} = \varphi_{A_E}$; поэтому, сложив почленно все равенства (10), получим

$$\varphi_C + \varphi_R + \varphi_L + \varphi_E = \varphi_{A_C} - \varphi_{B_E}.$$

Если цепь рис. 8 замкнута, то $\varphi_{A_C} = \varphi_{B_E}$. В этом случае, следовательно,

$$\varphi_C + \varphi_R + \varphi_L + \varphi_E = 0. \quad (13.1.11)$$

Это общее равенство вместе с выражениями (6)—(9) полностью определяет все процессы в цепи. Ниже мы, пользуясь этим равенством, будем рассматривать различные цепи, начиная с самых простых, состоящих только из д в у х элементов.

§ 2. Разряд емкости через сопротивление

Рассмотрим процесс, происходящий в цепи, состоящей из емкости C и сопротивления R (рис. 1). Потенциал точки A обозначим φ (пластину B будем считать заземленной). Вначале пусть $\varphi = \varphi_0$. Соответствующее количество электричества на пластине A будет $q_1 = C\varphi_0$.

Можно ли говорить о токе, идущем через емкость? В конденсаторе две пластины разделены изолятором (например, воздухом), так что в действительности электрон не может пройти через емкость, т. е. попасть из A в B. Однако если на пластину A попадает положительный заряд, то пластина B заряжается отрицательно, так что с пластины B по проводу уходит положительный заряд (ток также идет слева направо). Два амперметра A_1 и A_2, один из которых измеряет силу тока в проводе, присоединенном к пластине A, другой — в проводе, присоединенном к пластине B, дают одинаковые показания. Что именно (положительные заряды или электроны) проходит в различных частях электрической цепи, нас не интересует, так же как не интересует — те же электроны пройдут через A_2, которые ранее прошли через A_1, или другие. Поэтому везде в дальнейшем будем говорить о т о к е, и д у щ е м ч е р е з к о н д е н с а т о р, имея при этом в виду ток, проходящий по проводам, присоединенным к пластинам конденсатора. В электрической цепи о токе, идущем через конденсатор,

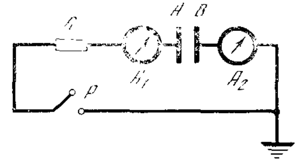


Рис. 13.2.1

можно говорить так же, как о токе, идущем через сопротивление или индуктивность; отличие заключается в другом виде связи между током и разностью потенциалов, что выражено формулами (1.9) и (1.9а).

При замыкании рубильника P (см. рис. 1) по сопротивлению R пойдет ток

$$j = \frac{1}{R}\varphi_R.$$

По формуле (1.11) $\varphi + \varphi_R = 0$, откуда $\varphi_R = -\varphi$; поэтому

$$j = -\frac{1}{R}\varphi. \quad (13.2.1)$$

Так как положительным здесь мы называем ток, текущий слева направо, то при $\varphi > 0$, как видно из формулы (1), ток отрицателен, течет справа налево, конденсатор разряжается[1]. Вспоминая, что $j = \frac{dq}{dt}$ (ток через конденсатор), а $q = C\varphi$, находим

$$j = C\frac{d\varphi}{dt}. \quad (13.2.2)$$

Сравнивая (1) и (2), находим

$$\frac{d\varphi}{dt} = -\frac{1}{RC}\varphi. \quad (13.2.3)$$

[1] Заметим, что во всех схемах (см. рис. 1.8 и далее), изображаемых в виде контура прямоугольника, мы говорим о направлении тока в в е р х н е й стороне прямоугольника; в нижней стороне, замыкающей цепь, ток имеет, очевидно, п р о т и в о п о л о ж н о е направление.

С подобным дифференциальным уравнением мы встречались в связи с задачей о радиоактивном распаде (см. гл. 8) и во многих других разделах этой книги. Если $\varphi=\varphi_0$ при $t=0$, то решение уравнения (3) таково:

$$\varphi(t)=\varphi_0 e^{-t/RC}. \qquad (13.2.4)$$

Отсюда

$$j(t)=-\frac{\varphi_0}{R}e^{-t/RC}.$$

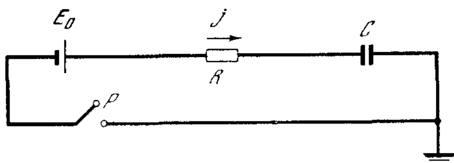


Рис. 13.2.2

Из формулы (3) видно, что величина RC имеет размерность *времени*. Проверим это:

$$[R]=\text{Ом}=\frac{\text{В}}{\text{А}}=\frac{\text{В}\cdot\text{с}}{\text{Кл}},$$

$$[C]=\frac{\text{Кл}}{\text{В}},$$

откуда

$$[RC]=\frac{\text{В}\cdot\text{с}}{\text{Кл}}\cdot\frac{\text{Кл}}{\text{В}}=\text{с}.$$

За время $t=RC$ заряд конденсатора q, а также сила тока j уменьшаются в e раз.

Процесс разрядки конденсатора легко проследить на опыте. В магазинах радиодеталей продаются конденсаторы емкостью $C=20$ мкФ$=20\cdot10^{-6}$ Ф и сопротивления $R=20$ МОм$=20\cdot10^6$ Ом. Для схемы с такими R и C получим $RC=400$ с — очень удобное для наблюдения время.

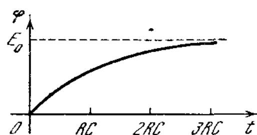


Рис. 13.2.3

Величина RC называется *постоянной времени* контура, состоящего из емкости и сопротивления (напомним, что в случае радиоактивного распада аналогичная величина называлась *средним временем* жизни).

Рассмотрим задачу з а р я д к и емкости через сопротивление. Соответствующая схема изображена на рис. 2. Если рубильник P замкнут, то согласно (1.11) $\varphi_E+\varphi_R+\varphi=0$, где φ — потенциал незаземленной пластины конденсатора. Так как $\varphi_E=-E_0$, $\varphi_R=Rj$, то $-E_0+Rj+\varphi=0$. Ток через конденсатор $j=\frac{dq}{dt}=C\frac{d\varphi}{dt}$; поэтому

$$-E_0+RC\frac{d\varphi}{dt}+\varphi=0,$$

или

$$\frac{d\varphi}{dt}=-\frac{1}{RC}(\varphi-E_0). \qquad (13.2.5)$$

Для того чтобы найти, как меняется φ с течением времени, удобно ввести новую переменную $z=\varphi-E_0$; тогда $dz=d\varphi$. При этом уравнение (5) примет следующий вид:

$$\frac{dz}{dt}=-\frac{z}{RC}.$$

Его решение таково:

$$z=z_0e^{-t/RC}, \qquad (13.2.6)$$

где z_0 — значение z в начальный момент времени.

Найдем решение для случая, когда в начальный момент времени конденсатор не заряжен: $\varphi=0$ при $t=0$. Тогда $z_0=-E_0$. Из (6) получаем $z=-E_0e^{-t/RC}$, т. е.

$$\varphi=z+E_0=-E_0e^{-t/RC}+E_0=$$
$$=E_0(1-e^{-t/RC}). \qquad (13.2.7)$$

График зависимости φ от t изображен на рис. 3. Кривая соответствует формуле (7), пунктирная горизонтальная прямая представляет собой то значение $\varphi=E_0$, к которому с течением времени приближается решение. Величина z имеет геометрический смысл расстояния по вертикали от кривой до пунктирной линии. Это расстояние с течением времени убывает по экспоненциальному закону.

За время, равное RC, заряд конденсатора достигает 63% своего конечного значения, за время $2RC$ — 86% и за время $3RC$ — 95% конечного значения.

Из формул (4) и (7) видно, что разрядка и зарядка конденсатора происходят тем быстрее, чем меньше сопротивление R и чем меньше емкость C.

Упражнения

13.2.1. В схеме рис. 1 $C=10^{-6}$ Ф и $R=10^7$ Ом; $R=10^8$ Ом; $R=10^9$ Ом. Для каждого из этих случаев определите, через сколько

времени ток, текущий в начальный момент через конденсатор, уменьшится на 10%; уменьшится вдвое.

13.2.2. Рассмотрите процесс выравнивания потенциала через сопротивление R при последовательном соединении двух конденсаторов C_1 и C_2, один из которых в начальный момент $t=0$ заряжен до разности потенциала $\varphi_{C_1}(0)=a$, а другой не заряжен вовсе, т. е. $\varphi_{C_2}(0)=0$ (рис. 4).

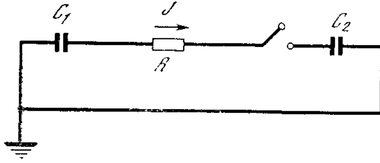


Рис. 13.2.4

13.2.3. Определить изменение постоянной времени контура в схеме рис. 1, если все линейные размеры схемы увеличить в n раз (конденсатор считать плоским). (Условия задачи надо понимать так: увеличиваются как размеры конденсатора, так и размеры сопротивления, но материалы, из которых изготовлены конденсатор и сопротивление, не меняются.) [Указание. Из физики известна формула для емкости плоского конденсатора: $C=\varepsilon\frac{S}{4\pi d}$, где S — площадь пластины конденсатора, d — расстояние между пластинами, ε — постоянная величина, зависящая от материала между пластинами (*диэлектрическая постоянная*). Величина проволочного сопротивления находится по формуле $R=\rho\frac{l}{\sigma}$, где l — длина, σ — площадь поперечного сечения, ρ — постоянная, зависящая от материала проволоки.]

§ 3. Колебания в цепи емкости с искровым промежутком

Типичная схема использования конденсатора показана на рис. 1. В цепь включен источник напряжения с ЭДС E и сопротивлением R (роль R может играть внутреннее сопротивление источника напряжения). Ниже находится *искровой промежуток*; при разности потенциалов меньше определенного значения φ_1 искровой промежуток является изолятором; при $\varphi=\varphi_1$ проскакивает искра, воздух между проводами накаляется до высокой температуры и становится хорошим проводником. Суммарное сопротивление подводящих проводов и накаленного воздуха обозначим через r. Величина r мала и остается малой до тех пор, пока идет ток, поддерживающий высокую температуру воздуха. При определенном малом значении тока j_2 воздух остывает и искровой промежуток снова становится изолятором. Этому значению тока соответствует разность потенциалов $\varphi_2=j_2r$. При этом $\varphi_1>\varphi_2$: нужно бо́льшее напряжение для того, чтобы зажечь искру, чем для того, чтобы поддержать ее горение.

На рис. 2 показана зависимость φ от t для такой схемы. На участке OA происходит зарядка емкости, ток не идет через искровой промежуток. В этом случае справедлива формула (2.6):

$$\varphi=E(1-e^{-t/(RC)}). \qquad (13.3.1)$$

В точке A в момент времени $t=t_A$ разность потенциалов достигает значе-

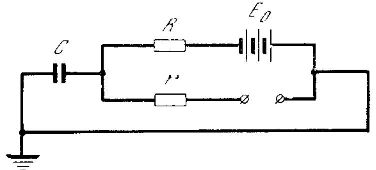


Рис. 13.3.1

ния φ_1, искровой промежуток начинает проводить ток, идет разрядка конденсатора. Так как при этом $R\gg r$, то током, идущим от источника напряжения, по сравнению с током, идущим через искровой промежуток, можно пре-

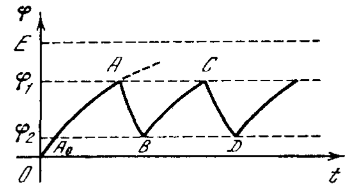


Рис. 13.3.2

небречь. Поэтому для φ получаем уравнение

$$\frac{d\varphi}{dt}=-\frac{\varphi}{rC},$$

причем $\varphi=\varphi_1$ при $t=t_A$. Отсюда находим

$$\varphi=\varphi_1 e^{-\frac{t-t_A}{rC}}. \qquad (13.3.2)$$

В момент времени $t=t_B$ (в точке B) $\varphi=\varphi_2$; при этом искровой промежуток опять становится изолятором, снова начинается зарядка (участок BC) и т. д.

Определим время t_B-t_A, в течение которого происходила разрядка емкости. Для этого воспользуемся тем,

что $\varphi=\varphi_2$ при $t=t_B$. Из уравнения (2), полагая в нем $\varphi=\varphi_2$, $t=t_B$, получим

$$\varphi_2=\varphi_1 e^{-\frac{t_B-t_A}{rC}},$$

откуда

$$t_B-t_A=rC\ln\frac{\varphi_1}{\varphi_2}.$$

На участке BC (зарядка) справедлива зависимость (1), сдвинутая по времени

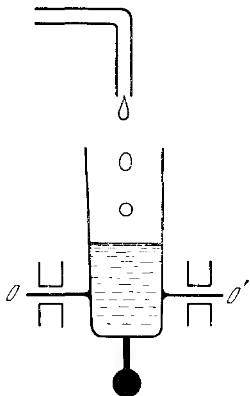


Рис. 13.3.3

на величину τ (на рис. 2 величина τ изображается отрезком A_0B). Поэтому

$$\varphi=E\left(1-e^{-\frac{t-\tau}{RC}}\right).$$

Полагая здесь $t=t_B$, получим

$$\varphi_2=E\left(1-e^{-\frac{t_B-\tau}{RC}}\right).$$

Аналогично, полагая $t=t_C$, находим

$$\varphi_1=E\left(1-e^{-\frac{t_C-\tau}{RC}}\right).$$

Из последних двух формул следует

$$\frac{E-\varphi_2}{E-\varphi_1}=e^{\frac{t_C-t_B}{RC}},\quad \text{или}$$

$$t_C-t_B=RC\ln\frac{E-\varphi_2}{E-\varphi_1}.$$

Полный период процесса (зарядка — разрядка) есть

$$T=t_C-t_A=(t_C-t_B)+(t_B-t_A)=$$
$$=RC\ln\frac{E-\varphi_2}{E-\varphi_1}+rC\ln\frac{\varphi_1}{\varphi_2}.$$

Обычно сопротивление в цепи источника напряжения R во много раз больше сопротивления искрового промежутка, поэтому время зарядки гораздо больше времени разрядки. Зато ток разрядки оказывается во много раз больше, чем ток зарядки, больше, чем тот максимальный ток, который можно было бы получить от источника напряжения (при наличии внутреннего сопротивления R_1 источник напряжения не дает ток больше, чем $\frac{E}{R_1}$). Схема на рис. 1 преобразует длительный малый ток, вызываемый источником напряжения, в сильный ток, который, однако, течет не все время, а в течение кратких промежутков времени (как принято говорить, наблюдаются «короткие импульсы» тока).

Действие схемы подобно системе, в которой маленькая струйка воды постепенно наполняет сосуд (рис. 3). Сосуд закреплен так, что, когда накопится достаточное количество воды, он опрокидывается, вода выливается, после чего сосуд снова принимает вертикальное положение, и процесс начинается сначала. На рисунке сосуд закреплен на горизонтальной оси OO', расположенной ниже середины сосуда. Внизу к сосуду прикреплен груз, так что центр тяжести пустого сосуда лежит ниже оси. Однако когда сосуд заполняется водой, то центр тяжести полного сосуда оказывается выше оси и сосуд опрокидывается.

Вернемся к схемам рис. 2.1 и 2.2. В таких схемах, состоящих из емкостей, сопротивлений и ЭДС, по истечении некоторого времени потенциалы практически становятся постоянными. Действительно, в схеме рис. 2.1 устанавливается $\varphi=0$, в схеме рис. 2.2 $\varphi=E_0$ (см. формулы (2.4) и (2.7))[1]. Совершенно другую картину наблюдаем в случае схемы с искровым промежутком. Здесь происходят незатухающие колебания величины φ (правда, они очень непохожи на те колебания, которые мы изучали раньше). Эти колебания связаны с особыми свойствами искрового промежутка, в частности с тем фактом, что до достижения определенного потенциала (так называемого *потенциала пробоя* φ_1) ток через искровой промежуток совсем не идет.

О свойствах разряда через воздух в искровом промежутке написано много книг; мы привели только минимум сведений, необходимых для понимания

[1] Ниже мы увидим, что это стремление потенциала к постоянному значению при $t\to\infty$ сохраняется и при наличии в схеме индуктивностей.

действия схемы рис. 1. Сказанного выше недостаточно, например, для ответа на простой вопрос: что произойдет, если соединить искровой промежуток с источником напряжения без конденсатора?

Действительно, если ток не идет, то на искровом промежутке будет напряжение E_0. Так как $E_0 > \varphi_1$, то должен произойти пробой. Но если бы произошел пробой, то сопротивление искрового промежутка стало бы малым, равным r. Тогда на искровом промежутке возникла бы разность потенциалов, равная $E_0 \frac{r}{r+R}$, и ток $j = \frac{E_0}{r+R}$. Если R велико, то ток j мал — меньше j_2, разность потенциалов на искровом промежутке мала — меньше φ_2. Но в таком случае воздух не нагревается, и сопротивление искрового промежутка не станет малой величиной r, а, значит, разность потенциалов будет велика, равна E. Мы получаем противоречие.

На самом деле в этих условиях получается электрический разряд другого типа, так называемый *тлеющий разряд* (малый ток без нагрева воздуха), а не искра с накаленным воздухом.

§ 4. Энергия конденсатора

Заряженный конденсатор обладает определенным запасом энергии; этот запас энергии может быть отдан очень быстро, если конденсатор разряжается на малое сопротивление.

Найдем запас энергии конденсатора емкости C, одна пластина которого заземлена, а другая имеет потенциал φ_0. Тогда количество электричества $q_0 = C\varphi_0$.

На первый взгляд энергия равна произведению $q_0\varphi_0$ $(=C\varphi_0^2)$. В действительности это выражение хоть и правильно передает порядок величины q_0, но все же не точно: оно в д в о е отличается от истинного.

Рассмотрим процесс зарядки конденсатора. В момент, когда его потенциал равен φ, а заряд q, добавка малого количества электричества dq увеличивает энергию на

$$dW = \varphi dq. \tag{13.4.1}$$

Существенно то, что по мере зарядки сам потенциал φ меняется, так как $\varphi = \frac{1}{C} q$. Подставляя это значение φ в (1), получим

$$dW = \frac{1}{C} q dq. \tag{13.4.2}$$

Проинтегрируем (2) от $q=0$ (незаряженный конденсатор) до $q=q_0$; получим

$$W(q_0) = \frac{1}{C}\int_0^{q_0} q dq = \frac{1}{2}\frac{q_0^2}{C} = \\ = \frac{1}{2}\varphi_0 q_0 = \frac{1}{2}C\varphi_0^2. \tag{13.4.3}$$

Таким образом, точный расчет заставляет нас приписать к произведению $q_0\varphi_0$ коэффициент 1/2.

Пусть теперь зарядка конденсатора производится от источника напряжения через сопротивление (см. рис. 2.2). Источник напряжения имеет постоянную ЭДС E_0. Поэтому, когда протекает количество электричества dq, источник напряжения совершает работу $E_0 dq$ (эта работа совершается за счет соответствующего уменьшения химической энергии источника напряжения). Следовательно, полная работа, совершаемая источником напряжения, равна $E_0 q_0$, где q_0 — полное протекшее количество электричества. При зарядке конденсатора процесс закончится, когда потенциал φ станет равным E_0. При этом источник напряжения произведет работу

$$E_0 q_0 = E_0 C E_0 = CE_0^2.$$

Каким запасом энергии будет обладать конденсатор? Это легко подсчитать по формуле (3):

$$W = \frac{1}{2} CE_0^2.$$

Куда же делась половина работы, совершенной источником напряжения? Она пошла на нагревание сопротивления R. Вспомним, что если через сопротивление протекало количество электричества dq, то при этом выделилась энергия

$$dA = \varphi_R dq, \tag{13.4.4}$$

где φ_R — разность потенциалов на концах сопротивления. Пользуясь тем, что $dq = j dt$, $j = \frac{\varphi_R}{R}$, мы можем преобразовать (4) к хорошо известному виду:

$$dA = \frac{(\varphi_R)^2}{R} dt = j^2 R dt.$$

Величина $j^2R=\frac{(\varphi_R)^2}{R}$ представляет собой количество энергии, выделяющейся на сопротивлении в единицу времени, т. е. *тепловую мощность*; ее размерность, как и следовало ожидать, Вт:

$[j^2R]=\text{А}^2\cdot\text{Ом}=\text{А}^2\cdot(\text{В}/\text{А})=\text{А}\cdot\text{В}=$
$=\text{А}\cdot(\text{Вт}/\text{А})=\text{Вт}.$

Зависимость $j(t)$ в случае зарядки конденсатора через сопротивление была найдена в § 2:

$$j(t)=\frac{E_0}{R}e^{-t/RC}.$$

Поэтому $dA=\frac{E_0^2}{R}e^{-2t/RC}dt.$

Энергия, выделившаяся за время T, есть

$$A(t)=\frac{E_0^2}{R}\int_0^T e^{-2t/RC}dt.$$

Отсюда

$$A(T)=-\frac{CE_0^2}{2}e^{-2t/RC}\Big|_0^T=$$
$$=\frac{1}{2}CE_0^2(1-e^{-2T/RC}). \qquad (13.4.5)$$

Мы знаем, что при неограниченном возрастании промежутка времени T по-

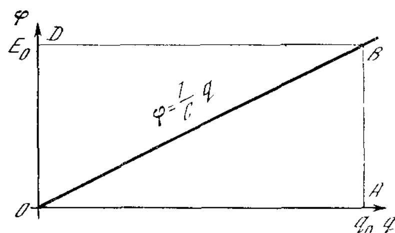


Рис. 13.4.1

тенциал φ неограниченно приближается к значению E_0. При этом, как видно из (5), величина A неограниченно приближается к $^1/_2CE_0^2$. Поэтому полная энергия, выделившаяся на сопротивлении,

$$A=\frac{1}{2}CE_0^2. \qquad (13.4.6)$$

Таким образом, расчет подтверждает, что при зарядке конденсатора половина энергии теряется на сопротивлении: КПД зарядки равен всего лишь 50%. Заметим, что если прямо присоединить провод от источника напряжения к конденсатору, то ничего не изменится: все равно КПД будет 50%, роль сопротивления R возьмет на себя внутреннее сопротивление источника напряжения, который будет при этом нагреваться. Из формулы (6) видно, что энергия, бесполезно теряемая на сопротивлении при зарядке конденсатора, не зависит от величины сопротивления R, а следовательно, не зависит от того, насколько быстро происходит зарядка.

Поскольку величина R не вошла в (6), то эту формулу можно получить, не вводя R в промежуточные преобразования. Действительно, для схемы рис. 2.2 $\varphi_E+\varphi_R+\varphi_C=0$, откуда $\varphi_R=-\varphi_E-\varphi_C=E_0-\frac{q}{C}$. Поэтому $dA=\left(E_0-\frac{q}{C}\right)dq$. Интегрируя это выражение от $q=0$ до $q=q_0=E_0C$, получим снова

$$A=\frac{1}{2}CE_0^2.$$

Последний вывод справедлив и для случая, когда сопротивление R меняется со временем; предыдущий же был справедлив лишь для $R=\text{const}$, так как только в этом случае можно применять формулы § 2.

Для того чтобы уменьшить потери при зарядке емкости, нужно было бы поступать так: сначала взять источник напряжения с малой ЭДС E_1 и зарядить конденсатор до потенциала E_1, затем первый источник напряжения отключить и присоединить второй источник с большей ЭДС E_2. Зарядив емкость до потенциала E_2, отключить второй источник и подключить третий с ЭДС E_3 и т. д. Получаемый при этом выигрыш легко представить графически: будем по оси абсцисс откладывать заряд конденсатора q, по оси ординат — его потенциал φ. Они связаны зависимостью $\varphi=\frac{1}{C}q$, которая изображается прямой линией (рис. 1). Энергия конденсатора равна площади треугольника OAB. Работа, произведенная источником напряжения, равна площади прямоугольника $OABD$. Энергия, потерянная на сопротивлении, равна площади треугольника ODB.

В случае зарядки конденсатора в несколько приемов сумма работ всех источников напряжения равна площади, заштрихованной на рис. 2. Предлагаем

читателю найти КПД для случая, когда процесс зарядки разбит на n этапов:

$$E_1=\frac{\varphi}{n},\quad E_2=\frac{2\varphi}{n},\quad E_3=\frac{3\varphi}{n},\ \ldots$$
$$\ldots,\quad E_n=\varphi.$$

Выше рассматривался конденсатор, одна пластина которого была заземлена, находилась при потенциале $\varphi_1=0$. При этом энергия конденсатора зависит от потенциала φ_2 второй пластины:

$$W=\frac{1}{2}C\varphi_2^2.$$

Если ни одна из пластин не заземлена, то энергия конденсатора зависит от разности потенциалов φ_C на пластинах:

$$W=\frac{1}{2}C\varphi_C^2.$$

Действительно, мы знаем, что заряд q на каждой пластине конденсатора зависит от разности потенциалов; при этом заряды на пластинах равны по величине и противоположны по знаку:

$$q_A=C\varphi_C,\quad q_B=-C\varphi_C=-q_A,$$
$$dq_A=-dq_B.$$

При расчете изменения энергии в процессе зарядки надо учесть изменение заряда обеих пластин. Пусть потенциал пластины A равен φ_1, потенциал пластины B равен φ_2, где $\varphi_1-\varphi_2=\varphi_C$. Тогда

$$dW=\varphi_1 dq_A+\varphi_2 dq_B=\varphi_1 dq_A-\varphi_2 dq_A=$$
$$=(\varphi_1-\varphi_2)\,dq_A=\varphi_C dq_A.$$

Так как $\varphi_C=\frac{q_A}{C}$, то

$$dW=\frac{1}{C}q_A dq_A. \qquad (13.4.7)$$

Интегрируя (7) от 0 до q_A, получим

$$W=\frac{q_A^2}{2C}=\frac{1}{2}C\varphi_C^2.$$

Зная выражение энергии заряженного конденсатора в зависимости от емкости, можно найти механические силы, действующие между пластинами конденсатора. Представим себе, что пластины конденсатора соединены механически с каким-то рычагом и емкость C зависит от положения рычага. Если положение рычага характеризуется значением координаты x, то емкость можно записать как $C(x)$. При определенном положении рычага x_0 емкость конденсатора $C(x_0)=C_0$. Если в этом положении конденсатор заряжен до потенциала φ_0, то заряд на пластинах $q_0=C_0\varphi_0$, энергия конденсатора

$$W=\frac{C_0\varphi_0^2}{2}=\frac{q_0^2}{2C_0}.$$

Отсоединим конденсатор от источника напряжения и будем передвигать рычаг. Тогда заряд сохраняется постоянным (потенциал меняется обратно

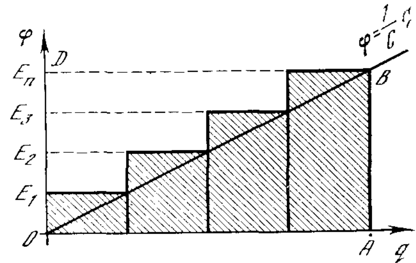


Рис. 13.4.2

пропорционально емкости), энергия меняется:

$$W(x)=\frac{q_0^2}{2C(x)}.$$

Электрическая энергия заряженного конденсатора подобна упругой энергии пружины. Увеличение $W(x)$ происходит в том случае, если внешняя сила, приложенная к рычагу, совершает работу. При этом внешняя сила преодолевает силы, действующие на рычаг со стороны пластины конденсатора. Напротив, при уменьшении $W(x)$ рычаг передвигается, совершая работу против внешних приложенных сил. Можно сделать вывод, что сила, действующая на рычаг со стороны пластин конденсатора, равна

$$F=-\frac{dW}{dx}=-\frac{d}{dx}\left[\frac{q_0^2}{2C(x)}\right]=$$
$$=\frac{q_0^2}{2[C(x)]^2}\frac{dC(x)}{dx}=\frac{\varphi^2(x)}{2}\frac{dC(x)}{dx}. \qquad (13.4.8)$$

Сила направлена в сторону увеличения емкости. Так, например, если конденсатор состоит из двух равных параллельных пластин, то емкость обратно пропорциональна расстоянию между пластинами. Значит, емкость увеличивается при сближении пластин. Действительно, при заряженном конденсаторе на пластинах находятся заряды противоположных знаков и пластины

притягивают друг друга тем сильнее, чем меньше расстояние между ними.

Формула (8) позволяет найти силу и в более сложных случаях, например в случае переменного конденсатора, в котором одна пластина может перемещаться в зазоре между двумя неподвижными.

Важно обратить внимание на то, что производную $\frac{dW}{dx}$ мы брали при заданном постоянном заряде q. Однако нельзя при нахождении силы по формуле $F = -\frac{dW}{dx}$ брать производную от $W = \frac{C(x)\varphi^2}{2}$, считая φ постоянным и учитывая, что C зависит от x. При этом мы получили бы неправильный знак силы. Действительно, если конденсатор отсоединен от источника напряжения, то φ непостоянно: $\varphi = \frac{q}{C}$, где $C = C(x)$. Если конденсатор присоединен к источнику напряжения, то при изменении емкости φ остается постоянным. Однако при этом меняется заряд q; значит, через источник напряжения течет ток, т. е. источник напряжения совершает (при увеличении C) работу, равную φdq. Следовательно, при постоянном φ при изменении емкости, применяя закон сохранения энергии, надо учитывать не только изменение энергии конденсатора и работу силы, но и работу, совершаемую источником напряжения.

§ 5. Цепь с индуктивностью

Рассмотрим цепь, состоящую из сопротивления R и индуктивности L (рис. 1). По формуле (1.11)

$$\varphi_R + \varphi_L = 0. \qquad (13.5.1)$$

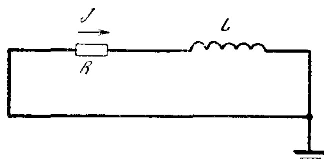


Рис. 13.5.1

Так как $\varphi_R = Rj$, а $\varphi_L = L\frac{dj}{dt}$, то, пользуясь (1), находим

$$Rj + L\frac{dj}{dt} = 0.$$

Таким образом, ток в цепи рис. 1 удовлетворяет уравнению

$$\frac{dj}{dt} = -\frac{R}{L}j. \qquad (13.5.2)$$

Решение этого уравнения имеет вид

$$j(t) = j_0 e^{-\frac{R}{L}t}. \qquad (13.5.3)$$

Мы видим, что в цепи рис. 1 сила тока уменьшается по экспоненциальному закону. Уменьшение силы тока в e раз происходит за время

$$T = \frac{L}{R}.$$

Проверим размерность величины $\frac{L}{R}$. Величина L измеряется в Г, т. е. в В·с/А, R — в Ом, т. е. в В/А. Поэтому

$$\left[\frac{L}{R}\right] = \left(\frac{\text{В} \cdot \text{с}}{\text{А}}\right) : \left(\frac{\text{В}}{\text{А}}\right) = \text{с},$$

так что действительно $\frac{L}{R}$ имеет размерность времени. Величину $\frac{L}{R}$ будем называть *временем затухания*. В схеме рис. 1, где нет источника напряжения, ток с течением времени стремится к нулю. Вопрос о том, как в такой схеме создать начальное значение тока j_0, мы рассмотрим немного ниже.

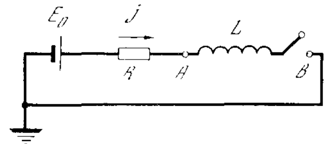


Рис. 13.5.2

Обратимся теперь к схеме, состоящей из источника с ЭДС, равной E_0, сопротивления R и индуктивности L (рис. 2). Из условия

$$\varphi_E + \varphi_R + \varphi_L = 0,$$

вспоминая, что $\varphi_E = -E_0$, находим

$$-E_0 + Rj + L\frac{dj}{dt} = 0. \qquad (13.5.4)$$

Перепишем это уравнение так:

$$\frac{dj}{dt} = \frac{R}{L}\left(\frac{E_0}{R} - j\right).$$

Это уравнение аналогично уравнению (2.5) и решается точно таким же приемом. Получаем

$$j(t) = \frac{E_0}{R} + Ae^{-\frac{Rt}{L}}, \qquad (13.5.5)$$

где значение A определяется из начального условия. Пусть в начальный момент времени $t = 0$ мы замкнули рубильник, до этого момента размыкающий цепь. При этом $j(0) = 0$, потому что при

разомкнутом рубильнике ток по цепи не шел. При таком условии находим $A = -\frac{E_0}{R}$ — и (5) принимает вид

$$j = \frac{E_0}{R}(1 - e^{-Rt/L}). \qquad (13.5.6)$$

С течением времени ток приближается к значению

$$j_\infty = \frac{E_0}{R}. \qquad (13.5.7)$$

Это значение тока не зависит от индуктивности L и просто получается из закона Ома в цепи с ЭДС E_0 и сопротивлением R. Однако такое значение тока устанавливается не сразу, а постепенно (рис. 3), и от индуктивности L зависит время установления тока: через время $\frac{L}{R}$ ток равен $0{,}63\ j_\infty$; через время $\frac{2L}{R}$ ток равен $0{,}86\ j_\infty$; через время $\frac{3L}{R}$ ток равен $0{,}95\ j_\infty$ и т. д; отношение L/R можно назвать *временем нарастания* тока.

Согласно основному уравнению (4) сумма разности потенциалов Rj на сопротивлении и $L\frac{dj}{dt}$ на индуктивности равна ЭДС E_0. Интересно проследить за каждым членом в отдельности. Они показаны на рис. 4. В начальный момент $j = 0$; $Rj = 0$, $E_0 = L\frac{dj}{dt}$; как говорят, «напряжение целиком садится на индуктивности». С течением времени ток приближается к постоянному значению, $\frac{dj}{dt}$ стремится к нулю, «напряжение садится на сопротивлении».

Интересно сопоставить решения, даваемые формулой (6) для одинакового E_0 и разных R и L. Пусть R_1 мало, а R_2 велико, L_1 мало, L_2 велико. При разных комбинациях R и L получим четыре зависимости тока от времени, изображенные на рис. 5. Конечный ток j_∞ зависит только от R, он одинаков для R_1, L_1 и для R_1, L_2; ток j_∞ одинаков также для пары кривых R_2, L_1 и R_2, L_2. Начальная скорость нарастания тока зависит только от индуктивности L и не зависит от сопротивления: кривые (R_1, L_1) и (R_2, L_1); (R_1, L_2) и (R_2, L_2) касаются в точке O (0, 0).

Из соображений размерности очевидно, что установившийся ток пропорционален начальной скорости нарастания тока и времени нарастания. При нашем определении времени нарастания отсюда уже получается совершенно правильная формула, без каких-либо добавочных коэффициентов. Действительно,

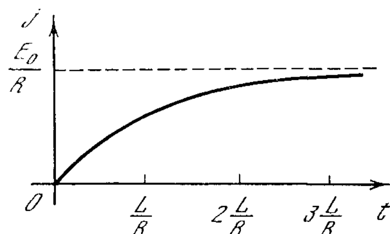


Рис. 13.5.3

начальная скорость нарастания тока $\left.\frac{dj}{dt}\right|_{t=0}$ равна $\frac{E_0}{L}$, время нарастания $T = \frac{L}{R}$, откуда и получается установившийся ток:

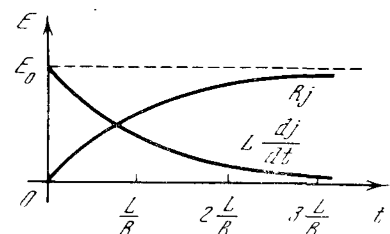


Рис. 13.5.4

$$j_\infty = T\left.\frac{dj}{dt}\right|_0 = \frac{L}{R}\frac{E_0}{L} = \frac{E_0}{R}.$$

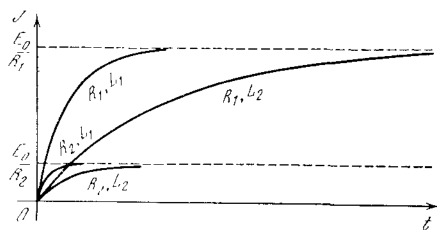


Рис. 13.5.5

Как осуществить начальный ток j_0 в схеме (см. рис. 1), с рассмотрения

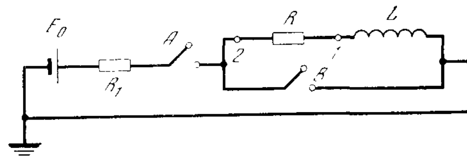


Рис. 13.5.6

которой мы начали? Для этого можно взять схему, приведенную на рис. 6. Сначала замкнем рубильник A при разомкнутом рубильнике B. Тогда в цепи

пойдет ток, который согласно формуле (7) вскоре достигнет значения $\frac{E_0}{R+R_1}$. Выберем E_0 так, что $\frac{E_0}{R+R_1}=j_0$. Дождемся установившегося состояния, когда в схеме с замкнутым A и разомкнутым B ток равен j_0. В этом состоянии замкнем рубильник B и разомкнем A. Получим схему рис. 1, причем в начальный момент времени (в момент замыкания B) в ней идет ток j_0. Потенциал в точке *1* до замыкания $\varphi_1=0$, так как в установившемся состоянии при постоянном j_0 падение напряжения на индуктивности L равно нулю. Потенциал в точке *2* до замыкания равен $\varphi_2=Rj_0$. При замыкании рубильника B точка *2* оказывается соединенной с «землей», потенциал в точке *2* $\varphi_2=0$. Соответственно перестраивается и потенциал в других точках цепи. В частности, в точке *1* потенциал теперь равен $\varphi_1=-Rj$.

§ 6. Размыкание цепи с индуктивностью

Выше был рассмотрен процесс установления тока в цепи рис. 5.2, состоящей из источника напряжения, сопротивления R, индуктивности L и рубильника. На рис. 5.3 показана кривая нараста-

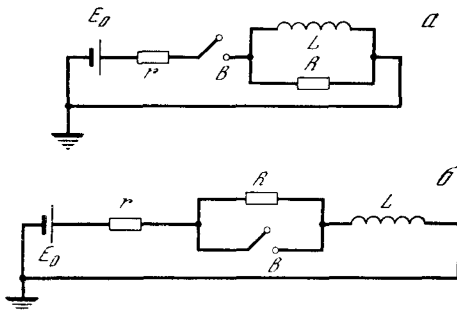


Рис. 13.6.1

ния тока, получающаяся при замыкании рубильника в момент $t=0$. С течением времени ток достигает значения $j_0=\frac{E_0}{R}$. Что произойдет, если теперь внезапно выключить рубильник B? Если ток прекратится за очень малое время τ, то производная тока по времени

$$\frac{dj}{dt}\approx\frac{j(t+\tau)-j(t)}{\tau}=-\frac{j_0}{\tau},$$

т. е. по абсолютной величине производная будет очень большой, поскольку мы считаем, что τ очень мало. При этом возникает очень большой по абсолютной величине отрицательный потенциал в точке A:

$$\varphi_A=L\frac{dj}{dt}\approx-L\frac{j_0}{\tau}.$$

Разность потенциалов на сопротивлении R, равная Rj, и ЭДС источника при размыкании меняются мало. Поэтому большая разность потенциалов, которая получается на индуктивности L, при размыкании почти целиком «садится» на рубильнике, т. е. разность потенциалов разомкнутых пластин рубильника оказывается очень большой, порядка $L\frac{j_0}{\tau}$, и может во много раз превосходить ЭДС источника тока E_0. При большой разности потенциалов происходит электрический пробой воздушного промежутка между разомкнутыми пластинами рубильника и между ними проскакивает искра.

Задача об изменении тока в цепи при размыкании рубильника оказывается очень сложной, так как сложны законы электрического разряда в воздухе между пластинами. В самом деле, до пробоя при $\varphi<\varphi_п$ ток не идет; однако, когда пробой произошел, сопротивление искры резко падает, идет большой ток при разности потенциалов, значительно меньшей $\varphi_п$. Отметим здесь только основной факт: в цепи с индуктивностью при размыкании возникают большие разности потенциалов; при замыкании такой цепи разность потенциалов нигде не превышает E_0 (ЭДС источника).

Проследить количественно явление кратковременного повышения разности потенциалов можно на примере двух схем рис. 1. Отличие этих схем от схемы рис. 5.2 заключается в том, что ток по индуктивности L может идти и при разомкнутом рубильнике B, так что размыкание происходит без искры. Однако если сопротивление R гораздо больше сопротивления r, то при размыкании возникает большая разность потенциалов на индуктивности.

Рассмотрим, например, схему рис. 1, *а*. Будем считать, что $R\gg r$. Если рубильник замкнут, то в произвольный момент времени ток в левой части цепи (r, E) равен сумме токов

в параллельно соединенных сопротивлении R и индуктивности L:

$$j_r = j_E = j_R + j_L.$$

При этом всегда $\varphi_L = \varphi_R$.

Пусть рубильник включен в момент $t = 0$. В этот момент весь ток идет через сопротивление R, так что $j_{r_0} = j_{R_0} = \frac{E_0}{R + r}$ по закону Ома. При этом

$$\varphi_{r0} = E_0 \frac{r}{r + R}, \quad \varphi_{R0} = E_0 \frac{R}{R + r};$$

следовательно,

$$\left.\frac{dj_L}{dt}\right|_{t=0} = \frac{\varphi_{R0}}{L} = \frac{E_0}{L} \frac{R}{R + r}.$$

По истечении достаточного времени после замыкания в цепи установится постоянная сила тока. В установившемся состоянии весь ток идет через индуктивность. Действительно, если ток j со временем не изменяется, то $\frac{dj}{dt} = 0$; поэтому $\varphi_L = 0$, а следовательно, $\varphi_R = 0$, откуда $j_R = 0$.

В установившемся состоянии $\varphi_{r\,\infty} = E_0$, $j_{r\,\infty} = j_{L\,\infty} = \frac{E_0}{r}$. Отсюда нетрудно оценить порядок времени установления тока τ_1:

$$j_{L\infty} - j_{L0} \approx \left.\frac{dj_L}{dt}\right|_{t=0} \tau_1,$$

или

$$\frac{E_0}{r} \approx \frac{E_0}{L} \frac{R}{R + r} \tau_1,$$

откуда

$$\tau_1 \approx \frac{L(R + r)}{rR} \approx \frac{L}{R}.$$

Рассмотрим теперь размыкание цепи, произведенное через время $t \gg \tau_1$ после замыкания, т. е. после того, как в цепи установилось постоянное значение тока $j_\infty = \frac{E_0}{r}$. Когда цепь разомкнута, $j_r = j_E = 0$ и $j_R + j_L = 0$, откуда $j_R = -j_L$, т. е. весь ток, проходящий через L, должен пройти через R в обратном направлении. По-прежнему, конечно, $\varphi_L = \varphi_R$. Поэтому $\varphi_R = Rj_R = -Rj_L$, или $\varphi_L = -Rj_L$. Так как $\varphi_L = L\frac{d\varphi_L}{dt}$, то

$$L\frac{dj_L}{dt} = -Rj_L.$$

Мы получили уравнение (5.2), что вполне естественно, так как правая часть схемы рис. 1, *а* (после размыкания рубильника) не отличается от схемы рис. 5.1.

Ток уменьшится в e раз за время $\tau_2 = \frac{L}{R}$. При этом $\tau_2 \ll \tau_1$, так как $R \gg\gg r$. В момент размыкания ток имеет значение $j_{L\infty} = \frac{E_0}{r}$. После того как размыкание произошло, но раньше, чем ток успел заметно уменьшиться, т. е. при t размыкания меньше τ_2, получим

$$\varphi_R = \varphi_L \approx -Rj_{L\infty} = -E_0 \frac{R}{r}.$$

Таким образом, при размыкании можно получить разность потенциалов,

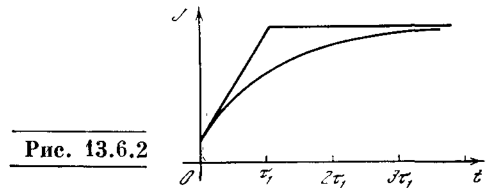


Рис. 13.6.2

во много раз превышающую ЭДС источника напряжения. Этот принцип широко используется в технике, в частности в системе зажигания двигателей внутреннего сгорания. Отметим, что эта большая разность потенциалов имеет место в течение весьма малого промежутка времени.

Мы привели приближенное рассмотрение задачи — фактически без использования производных и всего аппарата высшей математики. Точное рассмотрение задачи о замыкании рубильника в схеме рис. 1, *а* таково. Исходя из соотношений

$$\varphi_E + \varphi_r + \varphi_L = 0, \quad j = j_R + j_L, \quad \varphi_R = \varphi_L,$$

получим дифференциальное уравнение

$$\frac{dj_L}{dt} + \frac{rR}{(R + r)L} j_L = \frac{E_0 R}{(R + r)L}.$$

В начальный момент времени $t=0$ ток через индуктивность равен нулю, т. е. $j_L=0$ при $t=0$. Поэтому

$$j_L = \frac{E_0}{r}\left\{1 - \exp\left[-\frac{rR}{(R + r)L} t\right]\right\} = \frac{E_0}{r}\left(1 - e^{-\frac{t}{\tau_1}}\right). \qquad (13.6.1)$$

Ток в цепи

$$j = j_L + j_R = \frac{E_0}{r}\left(1 - e^{-\frac{t}{\tau_1}}\right) + \frac{E_0}{R+r} e^{-\frac{t}{\tau_1}}. \tag{13.6.2}$$

На рис. 2 приближенное решение соответствует ломаной, точное решение (2)— плавной кривой.

Рекомендуем читателю рассмотреть процесс изменения тока и разности потенциалов при замыкании и размыкании рубильника в схеме рис. 1, *б*. Полезно решить задачу дважды: один раз, составляя дифференциальное уравнение и отыскивая его решение в виде показательной функции, другой раз приближенно, подобно тому как была рассмотрена схема рис. 1, *а*.

§ 7. Энергия индуктивности

Выше мы видели, что в схеме, состоящей только из индуктивности L и сопротивления R, уже после того как эта схема отсоединена от источника напряжения, продолжает идти ток, постепенно затухающий с течением времени. При этом на сопротивлении в единицу времени выделяется тепло $Q = Rj^2$.

Откуда берется та электрическая энергия, которая в сопротивлении превращается в тепловую? Ее отдает индуктивность, обладающая определенным запасом энергии.

Найдем этот запас энергии, рассматривая простейшую схему рис. 5.1 и подсчитывая всю тепловую энергию, которая выделяется на сопротивлении R. В этой схеме пусть в начальный момент $t=0$ ток равен j_0. С течением времени ток убывает по закону

$$j(t) = j_0 e^{-Rt/L}.$$

Количество энергии, выделяющееся на сопротивлении R в единицу времени, т. е. *скорость выделения энергии* (ее размерность — Вт/с), есть мгновенная *тепловая мощность* h. Используя зависимость $j(t)$, приведенную выше, находим

$$h = Rj^2 = Rj_0^2 e^{-2Rt/L}. \tag{13.7.1}$$

Зная h, нетрудно найти полное количество тепла, выделившееся за время от начального момента $t=0$ до $t=\infty$ (до полного затухания тока). Для этого достаточно проинтегрировать (1) в интервале от $t=0$ до $t=\infty$. Мы получим

$$Q = \int_0^\infty Rj_0^2 e^{-2Rt/L} dt = Rj_0^2 \int_0^\infty e^{-2Rt/L} dt = \frac{Rj_0^2 L}{2R} = \frac{Lj_0^2}{2}. \tag{13.7.2}$$

Это тепло равно запасу энергии индуктивности, по которой идет ток j_0. Запас энергии не зависит от величины сопротивления R. Индуктивность L с током j_0 имеет определенный запас энергии, который в конце концов весь превращается в тепло независимо от величины сопротивления R. От R зависит только скорость превращения энергии в тепло, но не общее количество энергии.

Формулу (2) можно также получить, рассматривая процесс нарастания тока в индуктивности. Действительно, м о щ н о с т ь тока (работа в единицу времени) равна φj. Эта работа, совершаемая внешними источниками тока, тратится на увеличение энергии индуктивности W:

$$h = \frac{dW}{dt} = \varphi j. \tag{13.7.3}$$

Пользуясь тем, что $\varphi = L\frac{dj}{dt}$, получим из (3)

$$\frac{dW}{dt} = Lj\frac{dj}{dt} = \frac{1}{2}L\frac{d(j^2)}{dt}. \tag{13.7.4}$$

Будем считать, что $j=0$, $W=0$ при $t=0$ и $j=j_0$, $W=W_0$ при $t=t_0$. Тогда, интегрируя (4) от $t=0$ до $t=t_0$, получим

$$W_0 = \frac{1}{2}Lj_0^2.$$

Для конкретности можно представить себе схему рис. 5.2 ($\varphi = \varphi_A$) и для нее произвести подробный расчет нарастания энергии индуктивности. В установившемся режиме, когда ток достиг постоянного значения j_0, $\varphi_A = 0$, энергия индуктивности не меняется, однако источник ЭДС для поддержания постоянного тока j_0 должен продолжать расходовать энергию, которая выделяется в виде тепла на сопротивлении R.

Энергия индуктивности W пропорциональна квадрату силы тока, т. е. пропорциональна квадрату скорости движения электронов. Поэтому по внешнему виду W напоминает кинетическую энергию. Но есть ли W кинетическая энергия электронов?

Сравним порядок величины W и энергии электронов. Располагая медной проволокой длиной 100 м и диаметром 0,35 мм (так что сечение имеет порядок 10^{-7} м^2), можно намотать катушку, индуктивность которой составит 0,02 Г. При токе в 1 А в такой катушке $W=0{,}02\cdot 1^2\cdot 0{,}5=10^{-2}$ Дж. Найдем кинетическую энергию электронов.

Будем считать, что на каждый атом меди приходится один электрон, участвующий в прохождении тока (электрон проводимости). Относительная атомная масса меди равна примерно 63, так что на 63 г меди приходится $6\cdot 10^{23}$ электронов проводимости, или примерно 10^{22} электрон/г, или 10^{25} электрон/кг. Плотность меди около 8 г/см^3$=8\times 10^3$ кг/м^3; поэтому в 1 м^3 содержится примерно $n=8\cdot 10^{28}$ электронов проводимости. Можно представить себе кусок проволоки длиной vdt и сечением S слева от сечения O (рис. 1). Если скорость электронов[1] равна v, то через площадь S за время dt проходит $Snvdt$ электронов. За время dt электроны, находившиеся в сечении A, перейдут в сечение O; значит, за это время через O пройдут все электроны, которые находились в объеме между O и A, т. е. в объеме цилиндра с длиной образующей vdt и с основанием площади S.

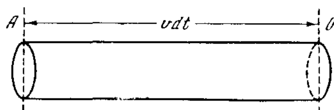


Рис. 13.7.1

Обозначим через e заряд одного электрона в кулонах, $e=-1{,}6\cdot 10^{-19}$ Кл. Количество электричества, которое переносят за время dt эти $Snvdt$ электронов, равно току в амперах, умноженному на время dt. Поэтому $Snvedt=jdt$, откуда $j=Snve$, или $v=\frac{j}{Sne}$.

Подставим в это соотношение $j=1$ А, $S=10^{-7}$ м^2, $n=8\cdot 10^{28}$ м$^{-3}$, $e=-1{,}6\cdot 10^{-19}$ Кл, получим

$$v=\frac{1}{10^{-7}\cdot 8\cdot 10^{28}\cdot 1{,}6\cdot 10^{-19}}\approx 8\cdot 10^{-4}.$$

При этом размерность

$$[v]=\text{А}/(\text{м}^2\cdot\text{м}^{-3}\cdot\text{Кл})=\text{А}\cdot\text{м}/(\text{А}\cdot\text{с})=\text{м/с}$$

— размерность скорости.

[1] Имеется в виду средняя скорость их движения в направлении тока, а не скорость хаотического теплового движения.

Найдем теперь кинетическую энергию электронов. Масса электрона $m=9\cdot 10^{-31}$ кг. Общее число электронов, движущихся в данной проволоке ($S=10^{-7}$ м^2, $l=10^2$ м), равно

$$10^2\cdot 10^{-7}\cdot 8\cdot 10^{28}\approx 10^{24}.$$

Кинетическая энергия одного электрона равна

$$\frac{1}{2}mv^2\approx\frac{1}{2}\cdot 9\cdot 10^{-31}\cdot 64\cdot 10^{-8}\approx$$
$$\approx 3\cdot 10^{-37}\ \text{Дж},$$

а общая кинетическая энергия всех электронов

$$T=n\frac{mv^2}{2}\approx 10^{24}\cdot 3\cdot 10^{-37}=$$
$$=3\cdot 10^{-13}\ \text{Дж}.$$

Таким образом, кинетическая энергия электронов составляет ничтожную долю энергии индуктивности, хотя она и зависит от силы тока по тому же закону (пропорциональна j^2), что и энергия индуктивности. Физическая энергия индуктивности есть энергия того магнитного поля, которое имеется в катушке, когда по ней течет ток.

Отметим сходство и различие между емкостью и индуктивностью. Как емкость, так и индуктивность могут служить резервуарами, хранилищами энергии. При помощи индуктивности, так же как и при помощи емкости, можно накопить электрическую энергию от слабого, маломощного первичного источника тока и затем быстро ее выделить в нужном месте в нужный момент.

Конденсатор можно заряжать малым током j_1 в течение большого времени t_1; быстро разряжая его через малое сопротивление, за малое время t_2 можно получить большой ток $j_2\approx\frac{j_1t_1}{t_2}$. Разность потенциалов конденсатора не превышает при этом ЭДС первичного источника. Конденсатор позволяет увеличить ток, но не напряжение.

Через индуктивность можно пропускать большой ток при малом напряжении (малой ЭДС) E_0 первичного источника. Для этого нужно только, чтобы сопротивление индуктивности и первичного источника тока было бы достаточно мало. При этом большой ток в индуктивности устанавливается не сразу, а за сравнительно большое время

t_3. При замыкании индуктивности на большое сопротивление удается получить большую разность потенциалов φ на малое время t_4, причем $\varphi \approx E_0 \frac{t_3}{t_4}$. Индуктивность позволяет увеличить напряжение, но не ток.

Существенное практическое отличие емкости от индуктивности заключается в том, что конденсатор, отсоединенный от источника тока, может очень долго (часами и даже сутками) сохранять запасенную в нем энергию. Время разрядки конденсатора равно RC, где C — емкость конденсатора, R — так называемое *сопротивление утечки*. Применяя хорошие изоляторы, удается получить огромные значения R, т. е. очень большое время разрядки. Индуктивность, выполненная в виде катушки и замкнутая накоротко, т. е. с минимальным возможным сопротивлением, при наличии в ней электрического тока сохраняет энергию лишь на протяжении долей секунды.

Время затухания тока имеет порядок $\frac{L}{R}$, но, даже применяя наилучшие проводники (медь, серебро), нельзя сделать $\frac{L}{R}$ больше нескольких секунд для катушки лабораторных размеров. Заметим, что при увеличении числа витков катушки в данном объеме за счет применения более тонкой проволоки увеличивается L — но увеличивается и R, и их отношение по порядку величины не изменяется. Поэтому в лабораторных условиях индуктивность удобно применять для повышения напряжения, но не для длительного хранения энергии.

В схемах, использующих емкости и индуктивности, удается накапливать энергию от батареи карманного фонаря. Такая батарея дает напряжение в несколько вольт при внутреннем сопротивлении в несколько ом, так что максимальная мощность ее порядка 1—2 Вт. При помощи упомянутых выше схем удается получить мощность в сотни киловатт. Однако такая мощность действует в течение времени порядка 10^{-6} с.

Выше было отмечено, что в индуктивности электрическая энергия быстро превращается в тепло за счет сопротивления. Это утверждение справедливо для катушек обычного «комнатного», лабораторного, масштаба и в случае обычных, нормальных, температур. Однако в двух крайних случаях это утверждение оказывается неверным.

1. При очень низкой температуре — от —260° С и ниже до абсолютного нуля (—273° С) — многие металлы (например, свинец, ртуть, но не медь) переходят в так называемое *сверхпроводящее* состояние. Их удельное сопротивление становится равным нулю.

Голландский ученый Каммерлинг-Оннес, открывший это явление в 1911 г., наблюдал в кольце из сверхпроводящего материала постоянный ток, который не ослабевал в течение многих суток. Наличие тока в таком кольце обнаруживается по магнитному полю этого тока.

Возможности практического применения сверхпроводников ограничиваются не только трудностью получения низкой температуры. Сильное магнитное поле переводит сверхпроводник в нормальное состояние (с конечным сопротивлением). Поэтому через сверхпроводник нельзя пропускать большие токи [2].

2. Соотношение между индуктивностью и сопротивлением и условия затухания тока очень сильно изменяются при увеличении всех размеров катушки и особенно при переходе к астрономическим явлениям (и астрономическим масштабам) [3].

Представим себе две геометрически подобные катушки, одна из которых в n раз больше другой. При этом число витков в обеих катушках одинаково. В большой катушке в n раз больше диаметр катушки, но во столько же раз больше и высота катушки и диаметр проволоки, из которой сделана ка-

---

[2] В 1961 г. открыт сплав редкого элемента ниобия с оловом, в котором ток до 100 000 А/см^2 и магнитное поле до 25 Тл (т. е. до 250 000 Гс) еще не разрушают сверхпроводимости. В настоящее время сверхпроводящие магниты широко применяются в науке и технике, в частности при конструировании ускорителей элементарных частиц. Сейчас уже можно также с определенностью утверждать, что мы близки и к осуществлению применения сверхпроводимости для передачи электроэнергии.

[3] Ср. с упр. 2.3: для контура, состоящего из емкости с сопротивлением, время разрядки не изменяется при изменении всех размеров.

тушка. Пусть катушки изготовлены из одинакового материала. Будем обозначать величины, относящиеся к меньшей катушке, индексом 1, а величины, относящиеся к большей катушке, индексом 2. Подсчитаем отношение сопротивлений катушек:

$$R_1=\rho\frac{l_1}{S_1},\quad R_2=\rho\frac{l_2}{S_2},$$

где ρ — удельное сопротивление материала катушки, l — длина проволоки, S — площадь ее поперечного сечения.

Геометрически ясно, что

$$l_2=nl_1,\quad S_2=n^2S_1;$$

следовательно,

$$R_2=\frac{1}{n}R_1$$

— сопротивление обратно пропорционально n, т. е. обратно пропорционально размеру.

Можно доказать, что индуктивность большой катушки ровно в n раз больше индуктивности малой катушки:

$$L_2=nL_1,$$

т. е. при увеличении линейных размеров катушки в n раз ее индуктивность увеличивается также в n раз. Время затухания тока τ имеет порядок $\frac{L}{R}$; следовательно,

$$\tau_1=\frac{L_1}{R_1},\quad \tau_2=\frac{L_2}{R_2}=n^2\frac{L_1}{R_1}=n^2\tau_1.$$

Таким образом, время затухания тока пропорционально к в а д р а т у размера катушки. Если бы земной шар состоял из меди, то время затухания тока в нем было бы порядка 10^{15}—10^{16} с, т. е. порядка 10^8 лет.

Проводимость ионизированных газов того же порядка, что и проводимость меди; поэтому в астрономических явлениях время затухания тока оказывается огромным. Это значит, что сопротивление, закон Ома не играют практически никакой роли в этих явлениях (напомним рис. 5.5: ток на начальном участке кривой зависит только от L, но не от R; в астрономии же мы всегда находимся на «начальном участке»).

Земной магнетизм представляет собой магнитное поле токов, протекающих в вязкой расплавленной массе центрального ядра Земли. Медленные движения расплавленной массы в этом магнитном поле поддерживают токи, подобно тому как движение якоря динамомашины в магнитном поле поддерживает ток в обмотке якоря и в обмотке электромагнита. То же относится и к магнитному полю Солнца. Теория успешно объясняет периодическое (полупериод $\tau\approx 11$ лет) изменение направления поля Солнца.

§ 8. Колебательный контур

Рассмотрим контур, состоящий из емкости C и индуктивности L (рис. 1). Пусть точка B контура заземлена. По формуле (1.11)

$$\varphi_C+\varphi_L=0,\quad \text{где}\quad \varphi_L=L\frac{dj}{dt},\quad \varphi_C=\frac{q}{C}.$$

Падение напряжения на емкости φ_C будем обозначать просто φ. Тогда

$$\varphi+L\frac{dj}{dt}=0. \qquad (13.8.1)$$

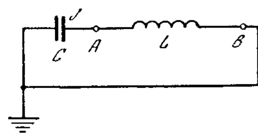


Рис. 13.8.1

Заметим, что $j=\frac{dq}{dt}$. Поэтому $\frac{dj}{dt}=\frac{d^2q}{dt^2}$. Так как $\frac{d^2q}{dt^2}=C\frac{d^2\varphi}{dt^2}$, то в силу (1) находим

$$\varphi+LC\frac{d^2\varphi}{dt^2}=0,$$

или

$$\frac{d^2\varphi}{dt^2}=-\frac{1}{LC}\varphi. \qquad (13.8.2)$$

С аналогичным уравнением мы встречались в гл. 10 при изучении механических колебаний. Там было установлено, что решением (2) являются функции $\varphi=A\sin\omega t$ и $\varphi=B\cos\omega t$ при любых A и B и надлежащим образом подобранном ω. Проверим это, например, для функции $\varphi=A\sin\omega t$; попутно мы определим частоту ω колебания. Подставим в (2) выбранное φ и $\frac{d^2\varphi}{dt^2}$. Мы получим

$$-ALC\omega^2\sin\omega t=-A\sin\omega t,$$

или после сокращения на $-A\sin\omega t$

$$LC\omega^2=1.$$

Отсюда находим

$$\omega = \frac{1}{\sqrt{LC}}. \qquad (13.8.3)$$

Следовательно, решением уравнения (2) являются функции, описывающие колебания с круговой частотой $\frac{1}{\sqrt{LC}}$; период таких колебаний равен

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{LC}. \qquad (13.8.4)$$

Проверим размерность (4):

$[C]$=Ф=Кл/В=А·с/В; $[L]$=Г= =В/(А/с)=В·с/А,

так что $[LC]=\text{c}^2$ и $T \sim \sqrt{LC}$ действительно имеет размерность с.

Обратимся к подробному рассмотрению решения уравнения (2). Решения $\varphi = A \sin \omega t$ и $\varphi = B \cos \omega t$ по существу не отличаются одно от другого, так как кривая синуса получается из кривой косинуса сдвигом по оси t; их можно объединить в одном решении

$$\varphi = B \cos(\omega t + \alpha) \qquad (13.8.5)$$

(ср. § 10.2). Амплитуда B колебаний и начальная фаза α (в § 10.2 мы ее обозначили буквой φ) могут быть любыми. При заданном $\varphi(t)$ найдем зависимость тока от времени:

$$j = C\frac{d\varphi}{dt} = -CB\omega \sin(\omega t + \alpha). \qquad (13.8.6)$$

Найдем энергию емкости и энергию индуктивности:

$$W_C = \frac{C\varphi^2}{2} = \frac{CB^2}{2}\cos^2(\omega t + \alpha),$$

$$W_L = \frac{Lj^2}{2} = \frac{LC^2B^2\omega^2}{2}\sin^2(\omega t + \alpha).$$

Подставляя сюда выражение (3) для ω, получим

$$W_L = \frac{CB^2}{2}\sin^2(\omega t + \alpha).$$

Полная энергия, как и следовало ожидать, от времени не зависит; действительно,

$$P = W_C + W_L = \frac{CB^2}{2}[\cos^2(\omega t + \alpha) + \sin^2(\omega t + \alpha)] = \frac{CB^2}{2}.$$

Таким образом, движение зарядов контура, составленного из емкости и индуктивности, похоже на движение массы, закрепленной на пружине. Энергию заряженного конденсатора можно уподобить упругой энергии пружины, которая максимальна в момент, когда масса находится в крайнем положении, на максимальном расстоянии от положения равновесия. Энергию индуктивности можно уподобить кинетической энергии движущейся массы. В момент, когда заряд емкости равен нулю, ток достигает максимального значения (по абсолютной величине); в этот момент энергия емкости равна нулю, а энергия индуктивности — полной энергии ($\cos^2(\omega t + \alpha) = 0$, $\sin^2(\omega t + \alpha) = 1$). Точно так же при колебаниях массы на пружине в момент, когда масса проходит через положение равновесия, потенциальная энергия равна нулю, а кинетическая энергия — полной энергии колебаний.

Назовем о б щ е й задачей задачу о нахождении потенциала в контуре при условии, что в начальный момент времени $t=0$ мы имеем $j=j_0$, $\varphi=\varphi_0$. Частные решения $\varphi = A \sin \omega t$ или $\varphi = B \cos \omega t$ здесь нам не подойдут — нам понадобится общее решение (5) с двумя (неопределенными) параметрами B и α, которыми мы можем распоряжаться.

Полагая в формулах (5) и (6) $t=0$, получим

$$\varphi(0) = B\cos\alpha = \varphi_0,$$

$$j(0) = -CB\omega \sin\alpha = j_0,$$

откуда найдем решение с заданными φ_0 и j_0:

$$B = \sqrt{\varphi_0^2 + \frac{j_0^2}{(C\omega)^2}},$$

$$\operatorname{tg}\alpha = -\frac{j_0}{C\omega\varphi_0}. \qquad (13.8.7)$$

Мы уже знаем, что при этом энергия постоянна и, следовательно, равна начальной энергии

$$\frac{C\varphi_0^2}{2} + \frac{Lj_0^2}{2}.$$

Выражение для энергии можно также записать как $C\frac{B^2}{2}$, что позволяет переписать первое из равенств (7):

$$B = \varphi_{\max} = \sqrt{\varphi_0^2 + \frac{L}{C}j_0^2}, \qquad (13.8.8)$$

откуда (и из (6)) также получаем

$$j_{\max} = \sqrt{j_0^2 + \frac{C}{L}\varphi_0^2} \qquad (13.8.9)$$

(заметьте, что в силу (4) $\omega=\sqrt{\frac{1}{LC}}$).

Схема для осуществления колебаний показана на рис. 2. В схеме имеется источник напряжения E_0. Если замкнуть A при разомкнутом B, то по истечении промежутка времени $\tau \gg RC$

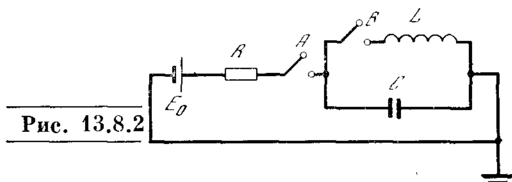


Рис. 13.8.2

после замыкания емкость будет заряжена до потенциала E_0. Разомкнем A и после этого в момент $t=0$ замкнем B. Тогда в цепи, состоящей из L и C, начнутся колебания с $\varphi=\varphi_0=E_0$; здесь $j=0$ в момент $t=0$. Отметим, что при этих колебаниях разность потенциалов на пластинах разомкнутого рубильника A будет периодически меняться от 0 до $2E_0$.

Возможен другой вариант возбуждений колебаний в схеме рис. 2. Сперва замыкаются оба рубильника A и B. Тогда в цепи устанавливается ток $j_0=\frac{E_0}{R}$. Затем в момент $t=0$ размыкается рубильник A. При этом в контуре, состоящем из L и C, начинаются колебания, причем в начальный момент $\varphi_0=0$ и $j_0=\frac{E_0}{R}$. При этих колебаниях максимальная амплитуда потенциала достигает

$$\varphi_{\max}=j_0\sqrt{\frac{L}{C}}=E_0\frac{1}{R}\sqrt{\frac{L}{C}}.$$

Напомним, что в схеме без емкости при размыкании цепи, содержащей индуктивность L, на рубильнике развивается тем большая разность потенциалов, чем больше сопротивление воздушного промежутка между пластинами рубильника. При размыкании такой цепи (без емкости) между пластинами рубильника обязательно возникает разряд в воздушном промежутке. При наличии емкости максимальная разность потенциалов между пластинами рубильника A не превышает определенного значения ($E_0+\varphi_{\max}$). Если это значение меньше того, которое необходимо для зажигания разряда в воздушном промежутке рубильника, то разряда не будет: говорят, что включение емкости C «гасит» разряд при размыкании цепи с индуктивностью. Отметим, что величина $\frac{1}{R}\sqrt{\frac{L}{C}}$ может быть больше единицы. Тогда, размыкая рубильник B через четверть периода после размыкания A, получим на емкости C потенциал более высокий, чем потенциал источника тока E_0.

Упражнения

13.8.1. Рассмотрите изменение потенциала со временем в цепи рис. 3. Определите наибольшее значение φ и время, когда достигается это наибольшее значение. (Будем считать, что рубильник A в схеме замыкают в момент $t=0$.)

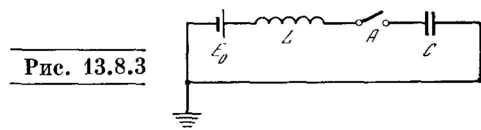


Рис. 13.8.3

13.8.2. В условиях упр. 1 найдите энергию емкости и энергию, отданную источником тока в момент, когда φ максимально.

§ 9. Затухающие колебания

Рассмотрим контур, в котором последовательно с индуктивностью включено сопротивление R (рис. 1). Будем считать, что R мало. Если R совсем не принимать во внимание, то мы получим схему рис. 8.1, рассмотренную

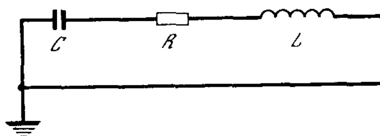


Рис. 13.9.1

в предыдущем параграфе. Если при $t=0$ было $\varphi=\varphi_0$, $j=0$, то согласно (8.7)

$$\varphi=\varphi_0\cos\omega t,$$

$$j=j_m\sin(\omega t+\pi)=-j_m\sin\omega t, \qquad (13.9.1)$$

где

$$j_m=j_{\max}=C\varphi_0\omega, \qquad \omega=\frac{1}{\sqrt{LC}}. \qquad (13.9.2)$$

При этом полная энергия $P=\frac{C\varphi_0^2}{2}$, что, пользуясь (2), можно записать еще и так:

$$P=\frac{Lj_m^2}{2}. \qquad (13.9.3)$$

При наличии сопротивления происходит превращение электрической энергии в тепловую. Тепловая мощность h равна

$$h = Rj^2 = Rj_m^2 \sin^2 \omega t = \frac{Rj_m^2}{2}(1 - \cos 2\omega t). \qquad (13.9.4)$$

Тепловая мощность при электрических колебаниях не остается постоянной: на протяжении каждого периода она дважды достигает максимума и дважды обращается в нуль (знак ее, конечно, не меняется — она всегда остается положительной). Найдем среднее значение h за период. Из (4) следует

$$\bar{h} = \frac{Rj_m^2}{2}(1 - \overline{\cos 2\omega t}).$$

Вспоминая, что среднее значение косинуса за период равно нулю, получаем

$$\bar{h} = \frac{Rj_m^2}{2}.$$

Выделение тепла на сопротивлении R может происходить только за счет уменьшения электрической энергии P. Поэтому

$$\frac{dP}{dt} = -h. \qquad (13.9.5)$$

Мы предположили, что R мало, значит, и h мало. Энергия колебаний убывает медленно, значительное изменение энергии заметно лишь по истечении нескольких периодов. Рассматривая промежутки времени, большие по сравнению с периодом колебаний T, заменим в правой части (5) h на $\bar{h}$:

$$\frac{dP}{dt} \approx -\bar{h} = -\frac{Rj_m^2}{2}. \qquad (13.9.6)$$

Поскольку энергия P медленно меняется, то из (3) следует, что и j_m есть медленно меняющаяся величина. Выразив j_m из (3), получим

$$j_m = \sqrt{\frac{2P}{L}}. \qquad (13.9.7)$$

Пользуясь (7), получаем из (6)

$$\frac{dP}{dt} = -\frac{R}{L}P$$

(знак $\approx$ мы заменили на $=$). Решение этого уравнения имеет вид

$$P = P_0 e^{-\frac{R}{L}t},$$

где P_0 — это значение P при $t=0$. Поэтому согласно (7)

$$j_m = \sqrt{\frac{2P_0}{L}} e^{-Rt/2L}.$$

Тогда

$$j = \sqrt{\frac{2P_0}{L}} e^{-Rt/2L} \sin(\omega t + \pi). \qquad (13.9.8)$$

Вспоминая, что $\varphi = \varphi_0 \cos \omega t$, а $\varphi_0 = \frac{j_m}{C\omega}$, получаем

$$\varphi = \frac{j_m}{C\omega} \cos \omega t = \frac{1}{C\omega} \sqrt{\frac{2P_0}{L}} e^{-Rt/2L} \cos \omega t. \qquad (13.9.9)$$

Формулы (8) и (9) показывают, что при наличии небольшого сопротивления электрические колебания затухают по экспоненциальному закону.

Выписанное выше решение получено при помощи приближенного расчета. Заметим, что в этом приближенном решении не удовлетворяется соотношение $j = C\frac{d\varphi}{dt}$, хотя оно выполняется тем точнее, чем меньше R. Постараемся теперь решить задачу точно. Для схемы рис. 1 имеем $\varphi + \varphi_R + \varphi_L = 0$, откуда

$$\varphi + Rj + L\frac{dj}{dt} = 0, \qquad (13.9.10)$$

причем $j = C\frac{d\varphi}{dt}$. Подставляя в (10) выражения для j и $\frac{dj}{dt}$, находим

$$LC\frac{d^2\varphi}{dt^2} = -\varphi - RC\frac{d\varphi}{dt}. \qquad (13.9.11)$$

Будем искать решение уравнения (11) в том же виде, который мы получили в приближенном рассмотрении вопроса, т. е. в виде

$$\varphi = Ae^{-\lambda t} \cos \omega t, \qquad (13.9.12)$$

где λ, ω, A — некоторые числа, которые нам еще требуется определить. Соответствующая процедура в точности совпадает с той, какая уже встречалась нам в теории колебаний (см. § 10.4; ср. решение (12) уравнения (11) с формулой (10.4.7)), так что мы могли бы здесь просто сослаться на результаты § 10.4. Но чтобы не затруднять читателя обращением к старому материалу и не заставлять его сопоставлять совершенно разные обозначения, повторим вкратце требуемый вывод.

Подставим в (11) выражение (12) для φ и следующие из (12) выражения для производных φ и сократим все члены на общий множитель $Ae^{-\lambda t}$. Мы получим

$$LC\lambda^2 \cos \omega t + 2LC\lambda\omega \sin \omega t - LC\omega^2 \cos \omega t = -\cos \omega t + RC\lambda \cos \omega t + RC\omega \sin \omega t.$$

Для того чтобы это равенство имело место при **любом** t, нужно, чтобы справа и слева совпадали коэффициенты при $\cos \omega t$ и при $\sin \omega t$:

$$LC\lambda^2 - LC\omega^2 = RC\lambda - 1; \qquad (13.9.13)$$

$$2LC\lambda\omega = RC\omega. \qquad (13.9.14)$$

Величину λ определяем из (14), после чего из (13) находим ω:

$$\lambda = \frac{R}{2L}, \qquad \omega = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}. \qquad (13.9.15)$$

Постоянная A из (11) не определилась; для того чтобы ее найти, надо использовать начальное условие (скажем, $\varphi = \varphi_0$ при $t = 0$). Наконец, зная $\varphi(t)$, легко найдем $j = C\frac{d\varphi}{dt}$:

$$j = -CAe^{-\lambda t}(\omega \sin \omega t + \lambda \cos \omega t). \qquad (13.9.16)$$

Сравнивая точное решение с приближенным, отметим, во-первых, что в приближенном рассмотрении задачи мы правильно определили число λ, характеризующее скорость затухания колебаний, однако не нашли зависимости частоты ω от сопротивления R; во-вторых, точная формула для тока несколько отличается от полученной в приближенном рассмотрении.

Точно таким же способом можно показать, что уравнение (11) допускает также решение

$$\varphi = Be^{-\lambda t} \sin \omega t \qquad (13.9.17)$$

с теми же ω и λ, что и ранее (см. 15)); этому решению отвечает сила тока

$$j = CBe^{-\lambda t}(\omega \cos \omega t - \lambda \sin \omega t). \qquad (13.9.18)$$

Сумма $\varphi = e^{-\lambda t}(A \cos \omega t + B \sin \omega t)$ решений (12) и (17) также является решением уравнения (11). Только при помощи этой суммы можно решить **общую** задачу: найти решение уравнения (11) с начальными условиями: $\varphi = \varphi_0$, $j = j_0$ при $t = 0$. Действительно, тогда для коэффициентов A и B получаем уравнения:

$$\varphi_0 = A, \quad j_0 = -CA\lambda + CB\omega,$$

откуда

$$A = \varphi_0, \quad B = \frac{C\lambda\varphi_0 + j_0}{C\omega}.$$

Упражнения

13.9.1. Найдите $j(t)$ для схемы рис. 1, если $C=1$, $L=1$, $R=0{,}1$; 0,5; 1 (мы считаем, что $\varphi=1$, $j=0$ при $t=0$).

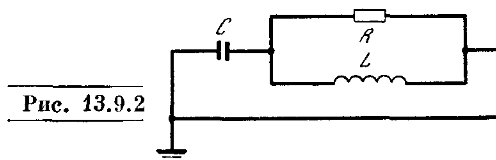


Рис. 13.9.2

13.9.2. Ответьте на тот же вопрос, если $\varphi=0$, $j=1$ при $t=0$.

13.9.3. Приближенным методом найдите скорость затухания колебаний λ в случае схемы рис. 2, считая, что R весьма велико.

§ 10*. Случай большого сопротивления

Рассматриваемый здесь случай **большого** сопротивления интересен главным образом с точки зрения математики и не связан с дальнейшим материалом. Поэтому в первом чтении этот параграф может быть пропущен.

Полученное в предыдущем параграфе решение уравнения (9.11) справедливо лишь для не слишком больших R. Действительно, из (9.15) видно, что если $R > 2\sqrt{\frac{L}{C}}$, то полученная нами формула для ω теряет смысл, так как в ней под корнем стоит отрицательное число. В этом случае уравнение (9.11) имеет решение совсем другого вида.

Будем искать решение (9.11) в виде $\varphi = Ae^{-\beta t}$ (соответственно $j = -AC\beta e^{-\beta t}$). Подставляя в (9.11) это выражение для φ и выражения для производных φ и сокращая все члены на $Ae^{-\beta t}$, получим

$$LC\beta^2 = -1 + RC\beta.$$

Это — квадратное уравнение относительно β. Решая его, найдем

$$\beta = \frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}. \qquad (13.10.1)$$

Подкоренное выражение в (1) отличается знаком от подкоренного выражения в (9.15) для ω. Следовательно, как раз в тех случаях, когда нельзя

найти ω, можно найти β, и наоборот. Формула (1) дает два различных значения β, поэтому можно составить два решения уравнения (9.11):

$$\varphi = Ae^{-\beta_1 t} \quad \text{и} \quad \varphi = Be^{-\beta_2 t}.$$

Решением будет и их сумма

$$\varphi = Ae^{-\beta_1 t} + Be^{-\beta_2 t}. \qquad (13.10.2)$$

Соответственно

$$j = -AC\beta_1 e^{-\beta_1 t} - BC\beta_2 e^{-\beta_2 t}. \qquad (13.10.3)$$

Если $\varphi = \varphi_0$, $j = j_0$ при $t = 0$, то, полагая в (2) и (3) $t = 0$, получим:

$$A + B = \varphi_0, \quad -AC\beta_1 - BC\beta_2 = j_0.$$

Из этой системы уравнений можно найти A и B.

Рассмотрим более подробно выражение для β. Пусть $R \gg 2\sqrt{\frac{L}{C}}$. Тогда $\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} = \frac{R}{2L}\sqrt{1 - \frac{4L}{R^2 C}}$ можно разложить по формуле бинома Ньютона (см. § 6.4). Ограничимся двумя членами разложения

$$\frac{R}{2L}\sqrt{1 - \frac{4L}{R^2C}} \approx \frac{R}{2L}\left(1 - \frac{1}{2}\frac{4L}{R^2C}\right) = \frac{R}{2L} - \frac{1}{RC}.$$

Поэтому

$$\beta_1 \approx \frac{R}{2L} + \frac{R}{2L} - \frac{1}{RC} = \frac{R}{L} - \frac{1}{RC} \approx \frac{R}{L},$$

так как R велико; аналогично

$$\beta_2 \approx \frac{R}{2L} - \frac{R}{2L} + \frac{1}{RC} = \frac{1}{RC}.$$

Эти значения β_1 и β_2 знакомы нам из § 1—5. Действительно, β_1 соответствует затуханию тока по закону $e^{-Rt/L}$, т. е. как в цепи, составленной только из индуктивности и сопротивления (см. § 5). Второй корень β_2 соответствует затуханию тока по закону $e^{-t/RC}$, т. е. в цепи, состоящей только из емкости и сопротивления (см. § 2).

С точки зрения математики интересен частный случай, когда подкоренное выражение в (1) точно равно нулю:

$$\frac{R^2}{4L^2} = \frac{1}{LC},$$

так что оба корня β_1 и β_2 совпадают. Мы получаем только одно решение уравнения (9.11). Однако для того чтобы решить задачу с начальными условиями $\varphi = \varphi_0$, $j = j_0$ при $t = 0$, нам надо иметь два решения.

Как найти второе решение? Предположим, что $\beta_1 \neq \beta_2$, но $\beta_1 - \beta_2$ — малая величина. Тогда мы имеем два решения: $e^{-\beta_1 t}$ и $e^{-\beta_2 t}$. Их разность также является решением. Запишем эту разность так:

$$e^{-\beta_1 t} - e^{-\beta_2 t} = e^{-\beta_2 t}\left[e^{(\beta_2 - \beta_1)t} - 1\right].$$

Так как $\beta_2 - \beta_1$ мало, то $e^{(\beta_2 - \beta_1)t} \approx 1 + (\beta_2 - \beta_1)t$ (в ряде Тейлора мы оставляем только два члена), откуда

$$e^{-\beta_1 t} - e^{-\beta_2 t} \approx e^{-\beta_2 t} t(\beta_2 - \beta_1).$$

Последнее выражение наводит на мысль, что в случае $\beta_2 = \beta_1 = \beta$ надо второе решение брать в виде $\varphi = Bte^{-\beta t}$. Подставляя это φ в уравнение (9.11) и учитывая, что $\beta = \frac{R}{2L}$, увидим, что уравнение действительно удовлетворяется. Итак, в случае $\beta_2 = \beta_1 = \beta$ надо брать φ в виде

$$\varphi = Ae^{-\beta t} + Bte^{-\beta t}.$$

Такое φ (и соответствующее ему j) позволяет решить задачу с любыми начальными φ_0 и j_0.

Упражнения

13.10.1. Найдите $\varphi(t)$, если $L = 1$, $C = 1$, $R = 2$; 6; 10 и $\varphi_0 = 1$, $j_0 = 0$ при $t = 0$.

13.10.2. Найдите $\varphi(t)$ для $L = 1$, $C = 1$, $R = 2$; 4 при условии, что $\varphi_0 = 1$, $j_0 = 1$ при $t = 0$.

§ 11. Переменный ток

Перейдем теперь к рассмотрению схем, в которых источник напряжения имеет ЭДС, *периодически меняющуюся* со временем с определенной (заданной) частотой ω. Эти задачи имеют большое значение для радиосхем. Частота переменного тока совершенно по-разному влияет на прохождение тока через индуктивность и емкость. Чем больше частота, тем быстрее меняется ток, тем «труднее» он проходит через индуктивность, тем больше разность потенциалов, создаваемая током данной силы. Напротив, на пластинах конденсатора при заданном токе разность потенциалов будет тем меньше, чем больше частота. При увеличении частоты уменьшается период, а следовательно, и время, в течение которого ток, сохраняя свое направление, заряжает конденсатор. Поэтому с увеличением частоты уменьшаются заряд конденсатора и разность потенциалов на его пластинах.

Мы уже отмечали в § 8, что движение зарядов в контуре, состоящем из индуктивности и емкости (контур L, C), можно уподобить обсуждавшемуся в гл. 10 колебанию тела, подвешенного на пружине: если при колебании тела периодически меняются его расстояние от начала координат и скорость, то в контуре периодически изменяются потенциал и сила тока.

Частота, с которой колеблется тело под действием упругой силы пружины (при отсутствии всяких других сил), называется собственной частотой колебаний. Подобно этому частота колебаний потенциала в контуре L, C называется *собственной частотой* этого контура.

Развивая эту аналогию дальше, можно предположить, что если контур включен в цепь переменного тока, т. е. периодически меняется подаваемый на контур потенциал, то будет иметь место явление *резонанса*. Оно состоит в том, что амплитуда колебаний максимальна при частоте тока ω, равной собственной частоте контура ω_0. Амплитуда очень резко возрастает при приближении разности $\omega - \omega_0$ к нулю. Явление резонанса на самом деле имеет место; оно будет рассмотрено в § 13.

Для каждого д в у х п о л ю с н и к а (см. рис. 1.7), включенного в цепь переменного тока, существует определенное соотношение между разностью потенциалов и силой тока. Найдем это соотношение сперва для наиболее простого случая отдельных элементов, с тем чтобы потом (в § 13 и 14) выяснить, как находить это соотношение для более сложных схем.

Будем рассматривать переменный ток определенной частоты ω (как и раньше, частота ω связана с периодом соотношением $\omega = \frac{2\pi}{T}$). Так, например, в СССР широко применяется 50-периодный ток:

$$T = \frac{1}{50}\ \text{с}, \quad \omega = 2\pi \cdot 50 \approx 314\ \text{с}^{-1}.$$

Представим себе схему, изображенную на рис. 1. В цепь включены амперметр A, показывающий силу тока j, и вольметр V, измеряющий напряжение (разность потенциалов). Допустим, что амперметр и вольтметр настолько безынерционные, быстродействующие, что они позволяют измерять м г н о в е н н о е значение тока в каждый момент и, следовательно, показания их изменяются с периодом, равным периоду тока. Практически такой опыт проводится с помощью *осциллографа*: так называемого *шлейфового осциллографа* с двумя шлейфами или с помощью *катодного осциллографа* с двумя лучами.

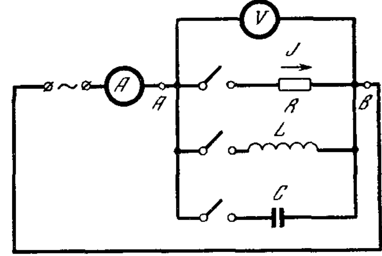


Рис. 13.11.1

Положительное направление тока показано стрелкой. Вольтметр V измеряет величину $\varphi = \varphi_A - \varphi_B$. Включая тот или иной рубильник, исследуем прохождение тока через сопротивление, индуктивность или емкость.

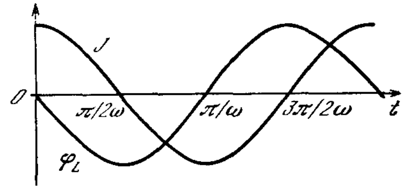


Рис. 13.11.2

Пусть ток меняется со временем по закону

$$j = j_0 \cos(\omega t + \alpha). \qquad (13.11.1)$$

Если этот ток идет по сопротивлению R, то по закону Ома

$$\varphi_R = Rj = Rj_0 \cos(\omega t + \alpha). \qquad (13.11.2)$$

Для общности запишем это равенство так:

$$\varphi_R(t) = \varphi_1 \cos(\omega t + \alpha_1),$$

где $\varphi_1 = Rj_0$, $\alpha_1 = \alpha$.

Пусть ток (1) идет через индуктивность L. Тогда

$$\varphi_L = L\frac{dj}{dt} = -L\omega j_0 \sin(\omega t + \alpha).$$

Положим

$$\varphi_L = \varphi_2 \cos(\omega t + \alpha_2). \qquad (13.11.3)$$

Тогда $\varphi_2 = L\omega j_0$, $\alpha_2 = \alpha + \frac{\pi}{2}$. Действительно, известно, что $\cos\left(\beta + \frac{\pi}{2}\right) = -\sin\beta$ при любом β; поэтому

$$\cos\left(\omega t + \alpha + \frac{\pi}{2}\right) = -\sin(\omega t + \alpha).$$

Таким образом, в случае переменного тока соотношение между амплитудой тока j и амплитудой напряжения φ_2 в индуктивности такое же, как в сопротивлении, равном $R_2 = L\omega$. При этом если L выражено в Г, ω — в обратных секундах (в $с^{-1}$), то R_2 выражено в Ом.

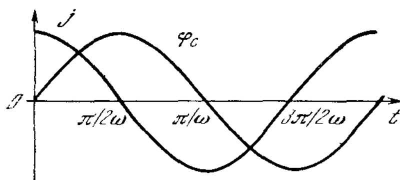


Рис. 13.11.3

Отличие индуктивности от сопротивления проявляется в том, что кривая напряжения на ней сдвинута относительно кривой тока на четверть периода (рис. 2). В этом можно убедиться и непосредственно по формуле

$$-\sin(\omega t + \alpha) = \cos\left(\omega t + \alpha + \frac{\pi}{2}\right) = \cos\left[\omega\left(t + \frac{\pi}{2\omega}\right) + \alpha\right].$$

Пусть функция $\cos(\omega t + \alpha)$, которой пропорционален ток, достигает какого-то определенного значения a в момент t_1:

$$\cos(\omega t_1 + \alpha) = a.$$

Функция $\cos\left(\omega t + \alpha + \frac{\pi}{2}\right)$, которой пропорционально напряжение на индуктивности, достигает того же значения a в некоторый другой момент времени t_2, так что

$$\cos\left(\omega t_2 + \alpha + \frac{\pi}{2}\right) = a = \cos(\omega t_1 + \alpha).$$

Поэтому $\omega t_2 + \frac{\pi}{2} = \omega t_1$, откуда $t_2 = t_1 - \frac{\pi}{2\omega} = t_1 - \frac{T}{4}$, т. е. напряжение опережает ток на четверть периода.

Разумеется, можно прибавить к t_1 любое целое число периодов, так что можно писать $t_2 = t_1 - \frac{T}{4} + T = t_1 + \frac{3}{4}T$ или $t_2 = t_1 + \frac{7}{4}T$. В формуле найден наименьший (по абсолютной величине) сдвиг по времени, который переводит кривую тока в кривую напряжения.

Рассмотрим случай емкости. Здесь $j = C\frac{d\varphi_C}{dt}$, поэтому

$$\varphi_C = \frac{1}{C}\int j\,dt = \frac{1}{C}\int j_0\cos(\omega t + \alpha)\,dt = \frac{j_0}{C\omega}\sin(\omega t + \alpha) \qquad (13.11.4)$$

(постоянная интегрирования равна $\bar{\varphi}_C$, но для переменного тока всегда $\bar{\varphi}_C = 0$). Записав φ_C в виде $\varphi_C = \varphi_3\cos(\omega t + \alpha_3)$, получим:

$$\varphi_3 = \frac{1}{C\omega}j_0, \quad \alpha_3 = \alpha - \frac{\pi}{2}.$$

Следовательно, в цепи переменного тока соотношение между амплитудой тока и амплитудой напряжения на емкости такое же, как на сопротивлении, равном $R_3 = \frac{1}{C\omega}$. Выражая емкость в Ф и частоту в $с^{-1}$, получим R в Ом.

Кривая напряжения в емкости сдвинута относительно кривой тока вперед на четверть периода (рис. 3). Таким образом, кривая напряжения φ_C на емкости сдвинута в сторону, противоположную кривой φ_L.

При данном одинаковом токе φ_L и φ_C имеют противоположные знаки. Если же совместить кривые φ_L и φ_C, то окажется, что ток, идущий через емкость, и ток, идущий через индуктивность, имеют противоположные знаки. Действительно, все формулы, выражающие изменение φ в зависимости от времени, легко преобразовать в формулы, выражающие процесс изменения j:

$$\varphi = \varphi_0\cos(\omega t + \alpha),$$

$$\varphi_R = Rj_0\cos(\omega t + \alpha),$$

$$\varphi_L = L\omega j_0\cos\left(\omega t + \alpha + \frac{\pi}{2}\right) = -L\omega j_0\sin(\omega t + \alpha),$$

$$\varphi_C = \frac{1}{C\omega}j_0\cos\left(\omega t + \alpha - \frac{\pi}{2}\right) = \frac{1}{C\omega}j_0\sin(\omega t + \alpha)$$

и

$$j = j_0 \cos(\omega t + \alpha),$$

$$j_R = \frac{\varphi_0}{R} \cos(\omega t + \alpha),$$

$$j_L = \frac{\varphi_0}{L\omega} \cos\left(\omega t + \alpha - \frac{\pi}{2}\right) = \frac{\varphi_0}{L\omega} \sin(\omega t + \alpha),$$

$$j_C = \varphi_0 C\omega \cos\left(\omega t + \alpha + \frac{\pi}{2}\right) = -\varphi_0 C\omega \sin(\omega t + \alpha).$$

Противоположный сдвиг фазы и противоположные знаки в формулах, относящихся к индуктивности и емкости, имеют решающее значение при рассмотрении L и C, включенных вместе.

В опытах с переменным током часто пользуются *однолучевым катодным осциллографом*. При этом на одну пару отклоняющихся пластин (отклонение по оси x) подают напряжение, пропорциональное силе тока. На другую пару пластин (отклонение по оси y) подают напряжение, пропорциональное φ. Луч движется по линии, уравнение которой имеет вид $x=aj$, $y=b\varphi$; коэффициенты a и b зависят от чувствительности осциллографа. Так как j и φ суть периодические функции времени, то луч прочерчивает на экране все время одну и ту же линию. При частоте 50 периодов в секунду глаз не замечает движения луча, а видит сплошную светящуюся линию.

Если на пластинки вертикального отклонения осциллографа подается разность потенциалов φ_R с сопротивления, то луч вычерчивает прямую. В самом деле,

$$x = aj = aj_0 \cos(\omega t + \alpha),$$

$$y = b\varphi_R = bRj_0 \cos(\omega t + \alpha),$$

и, исключая отсюда t, найдем

$$y = \frac{bR}{a} x.$$

Если на эти пластинки подается разность потенциалов φ_C с емкости, то получается эллипс:

$$x = aj_0 \cos(\omega t + \alpha),$$

$$y = b \frac{1}{C\omega} j_0 \sin(\omega t + \alpha),$$

откуда

$$\left(\frac{x}{aj_0}\right)^2 + \left(\frac{y}{bj_0/C\omega}\right)^2 = \cos^2(\omega t + \alpha) + \sin^2(\omega t + \alpha) = 1.$$

Точно так же эллипс получается и при включении φ_L. При включении $\varphi_R + \varphi_L$ или $\varphi_R + \varphi_C$ оси симметрии эллипса уже не совпадают с осями x и y. Таким образом, по форме осциллограммы можно судить, что включено в цепь (C, L или R), т. е. чем «начинен ящик», изображенный на рис. 1.7.

§ 12. Средние величины. Мощность и сдвиг фазы

В предыдущем параграфе ток и напряжение считались переменными величинами, меняющимися с течением времени (функции времени). Однако во многих вопросах достаточно знать (постоянные) средние значения этих величин (ср. § 7.7).

Как простейший пример рассмотрим нагревательный прибор с сопротивлением R. Мы знаем, что в цепи постоянного тока мощность (т. е. количество энергии, выделяющейся в единицу времени) равна $h = \varphi j = Rj^2 = \frac{\varphi^2}{R}$. В цепи переменного тока мощность тока меняется: (мгновенная) мощность в момент t равна

$$h(t) = \varphi(t)\, j(t) = R\,[j(t)]^2 = \frac{[\varphi(t)]^2}{R}.$$

Поэтому (см. (11.1))

$$h(t) = Rj_0^2 \cos^2(\omega t + \alpha). \qquad (13.12.1)$$

На протяжении одного периода $h(t)$ дважды обращается в нуль и дважды достигает максимального значения Rj_0^2. Но при рассмотрении нагревательных приборов нас обычно интересует количество тепла, выделившееся за время t, намного превышающее период T переменного тока (в СССР обычно равный 0,02 с). Поэтому нам достаточно знать среднее значение мощности h за большой промежуток времени t. В силу (1)

$$\bar{h} = \overline{Rj_0^2 \cos^2(\omega t + \alpha)} = Rj_0^2\, \overline{\cos^2(\omega t + \alpha)} \approx \frac{Rj_0^2}{2}, \qquad (13.12.2)$$

ибо, как мы уже неоднократно указывали в этой книге, $\overline{\cos^2(\omega t + \alpha)} = \frac{1}{2}$

(ср. § 7.7 или 10.4); при этом равенство (2) является приближенным, и оно тем точнее, чем больше t.

Среднее значение силы переменного тока $\bar{j}$ принято определять как *силу постоянного тока, выделяющего на сопротивлении* R *такую же мощность, как наш переменный ток*[1]:

$$R\bar{j}^2 = \frac{Rj_0^2}{2}. \qquad (13.12.3)$$

Отсюда

$$\bar{j} = \frac{1}{\sqrt{2}} j_0 \approx 0{,}71 j_0. \qquad (13.12.4)$$

Точно так же среднее значение напряжения $\bar{\varphi}$ определяется из условия

$$\bar{h} = \frac{\bar{\varphi}^2}{R} = \frac{\varphi_0^2}{2R},$$

откуда

$$\bar{\varphi} = \sqrt{\overline{\varphi^2}} = \frac{1}{\sqrt{2}} \varphi_0 \approx 0{,}71\varphi_0. \qquad (13.12.5)$$

Приборы, измеряющие переменный ток, — амперметры, вольтметры — градуированы так, что стрелки показывают на шкале именно с р е д н и е з н а ч е н и я $\bar{j}$ и $\bar{\varphi}$.

Из (4) и (5) следует, что максимальные значения силы тока и напряжения, достигаемые в цепи переменного тока, в $\sqrt{2} \approx 1{,}41$ раз больше средних. Так, например, в цепи со средним напряжением 220 В максимальное мгновенное напряжение достигает ± 310 В.

Из соотношений (2), (4), (5) и формулы $\varphi_0 = Rj_0$ следует, что $\bar{\varphi} = R\bar{j}$ и $\bar{h} = \bar{\varphi}\bar{j}$, так что закон Ома и связь между мощностью, током и напряжением на сопротивлении остаются справедливыми и для средних значений.

При прохождении переменного тока через емкость и индуктивность мы встретились с тем, что сила тока и напряжение на них меняются по законам, графики которых сдвинуты один относительно другого, хотя частоты изменения этих величин одинаковы. Рассмотрим мощность в общем случае произвольных начальных фаз величин j и φ (произвольного сдвига одного графика относительно другого). Пусть

$$j = j_0 \cos(\omega t + \beta),$$

$$\varphi = \varphi_0 \cos(\omega t + \beta + \alpha);$$

тогда

$$h(t) = j_0\varphi_0 \cos(\omega t + \beta) \cos(\omega t + \beta + \alpha).$$

Но

$$\cos(\omega t + \beta + \alpha) = \cos(\omega t + \beta)\cos\alpha - \\ - \sin(\omega t + \beta)\sin\alpha,$$

так что

$$\cos(\omega t + \beta)\cos(\omega t + \beta + \alpha) = \\ = \cos\alpha\cos^2(\omega t + \beta) - \\ - \sin\alpha\cos(\omega t + \beta)\sin(\omega t + \beta) = \\ = \cos\alpha\cos^2(\omega t + \beta) - \\ - \sin\alpha \cdot {}^1/_2 \sin(2\omega t + 2\beta).$$

А так как

$$\overline{\cos^2(\omega t + \beta)} = {}^1/_2, \quad \overline{\sin(2\omega t + 2\beta)} = 0,$$

то

$$\bar{h} = j_0\varphi_0 \cos\alpha \cdot {}^1/_2 = \bar{j}\bar{\varphi}\cos\alpha.$$

Таким образом, среднее значение мощности в общем случае, при наличии сдвига фаз на α, пропорционально $\cos\alpha$. В частном случае сопротивления $\alpha = 0$, $\cos\alpha = 1$, — и мы снова приходим к формуле (2).

В случае емкости $\alpha = -\frac{\pi}{2}$, $\cos\alpha = 0$; в случае индуктивности $\alpha = +\frac{\pi}{2}$, $\cos\alpha = 0$. Таким образом, в обоих случаях с р е д н я я м о щ н о с т ь р а в н а н у л ю. Этот результат вполне понятен физически. В емкости и индуктивности электрическая энергия не превращается в тепло, а только может запасаться. В цепи переменного тока на протяжении одной части периода емкость забирает из цепи электрическую энергию и запасает ее; во время другой части периода емкость отдает обратно свою энергию, работает на цепь. То же самое относится и

---

[1] Ясно, что если величина j задается формулой (11.1), то вводимое по формулам § 7.7 среднее значение $\bar{j} = 0$; поэтому целесообразно именно сформулированное определение среднего значения $\bar{j} \equiv \sqrt{\overline{j^2}}$ (эта величина в математике называется *средним квадратичным*); корень квадратный необходим здесь для того, чтобы $\bar{j}$ имело размерность силы тока, а операция возведения в квадрат позволяет избежать взаимной компенсации положительных и отрицательных значений тока; знак $\equiv$ в записи $\bar{j} \equiv \sqrt{\overline{j^2}}$ означает «равно по определению» (определяется здесь по-новому величина $\bar{j}$, число $\overline{j^2}$ имеет обычный смысл).

к индуктивности, включенной в цепь переменного тока.

Обычный трансформатор, к которому не подключено никакой нагрузки, представляет собой чистую индуктивность (если пренебречь небольшими потерями в проводах). Через такой трансформатор течет ток с амплитудой $j = \varphi/L\omega$. Однако мощность из сети, как уже было сказано, не отбирается благодаря тому, что фаза $\alpha = \frac{\pi}{2}$, $\cos\alpha = 0$. Интересно отметить, что счетчики израсходованной электроэнергии устроены так, чтобы измерять именно величину $j\varphi\cos\alpha$. По этой причине трансформатор, включенный без нагрузки, практически ничего не прибавляет к Вашему счету за электроэнергию, а только увеличивает полный ток, текущий по проводам.

Если в сеть параллельно включено много индуктивностей (трансформаторы, электромоторы без нагрузки и т. д.), то полный ток может стать довольно большим, и тогда станут заметными потери в электропроводке. Этот эффект становится важным для электросетей в масштабе города. Законы протекания тока по схемам с индуктивностями и емкостями показывают, как можно бороться с этими потерями.

§ 13. Колебательный контур в цепи переменного тока. Резонанс напряжений

Рассмотрим теперь включенные последовательно в цепь переменного тока сопротивление, индуктивность и емкость (рис. 1). Очевидно, что в этой системе ток, идущий через L, R, C, одинаков. Запишем его в виде

$$j = j_0 \cos(\omega t + \alpha). \qquad (13.13.1)$$

Разность потенциалов в цепи $\varphi = \varphi_1 - \varphi_4 = \varphi_R + \varphi_L + \varphi_C$.

Припоминая формулы (11.2)—(11.4), получим

$$\varphi = Rj_0 \cos(\omega t + \alpha) - L\omega j_0 \sin(\omega t + \alpha) + \frac{j_0}{C\omega}\sin(\omega t + \alpha) = Rj_0\cos(\omega t + \alpha) + j_0\left(\frac{1}{C\omega} - L\omega\right)\sin(\omega t + \alpha). \qquad (13.13.2)$$

Из (2) видно, что разности потенциалов на индуктивности и на емкости имеют разные знаки, благодаря чему коэффициент при $\sin(\omega t + \alpha)$ представляется в виде разности двух чисел. Запишем φ в виде

$$\varphi = b\cos(\omega t + \alpha + \beta), \qquad (13.13.3)$$

где b есть амплитуда разности потенциалов, т. е. наибольшее значение разности потенциалов (напряжения). Для того чтобы найти b, перепишем (3) так:

$$\varphi = b\cos\beta\cos(\omega t + \alpha) - b\sin\beta\sin(\omega t + \alpha). \qquad (13.13.3a)$$

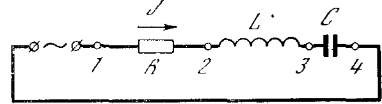


Рис. 13.13.1

Из сравнения (3а) с (2) вытекает:

$$b\cos\beta = Rj_0,$$
$$b\sin\beta = j_0\left(L\omega - \frac{1}{C\omega}\right). \qquad (13.13.4)$$

Возводя равенства (4) в квадрат, складывая результаты и извлекая из суммы квадратный корень, найдем

$$b = j_0\sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2}. \qquad (13.13.5)$$

Из (5) видно, что при данном значении амплитуды тока амплитуда напряжения минимальна при

$$L\omega = \frac{1}{C\omega}. \qquad (13.13.6)$$

Так как в силу (5)

$$j_0 = b\Bigg/\sqrt{R^2 + \left(L\omega - \frac{1}{C\omega}\right)^2},$$

то (6) есть также условие максимальности амплитуды тока j_0 при данном значении амплитуды напряжения b. Это условие можно записать еще в виде $\omega = \frac{1}{\sqrt{LC}}$, доказывающем совпадение ω с собственной частотой контура с данными L и C. Поэтому (6) есть условие *резонанса*, условие совпадения собственной частоты контура с частотой того переменного тока, который мы подаем. Отметим, что при резонансе напряжение в цепи равно

$$\varphi = Rj_0\cos(\omega t + \alpha). \qquad (13.13.7)$$

Пользуясь (1), находим, что в этом случае

$$\bar{\varphi}=R\bar{j}. \tag{13.13.8}$$

Перейдем к средним значениям. Средние значения величин тока и разности потенциалов определяются в соответствии с формулами (12.4) и (12.5). Из формулы $\varphi_L=L\omega j_0 \sin(\omega t+\alpha)$ следует (см. сноску 1 в § 12)

$$\bar{\varphi}_L=L\omega\bar{j}=\frac{L\omega}{R}\bar{\varphi}. \tag{13.13.9}$$

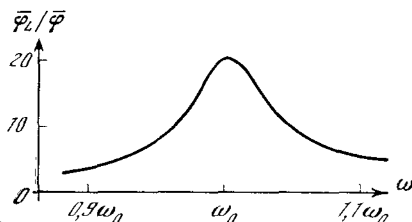


Рис. 13.13.2

Аналогично из формулы $\varphi_C=\frac{1}{C\omega}j_0\times\times\sin(\omega t+\alpha)$ имеем

$$\bar{\varphi}_C=\frac{1}{C\omega}\bar{j}=\frac{\bar{\varphi}}{C\omega R}. \tag{13.13.10}$$

Формулы (7)—(10) имеют место только в случае резонанса, т. е. при $\omega_0=\frac{1}{\sqrt{LC}}$. Подставляя это значение ω в (9) и (10), получим

$$\bar{\varphi}_L=\bar{\varphi}_C=\frac{1}{R}\sqrt{\frac{L}{C}}\bar{\varphi}.$$

Поэтому в случае резонанса напряжение на индуктивности и емкости тем больше, чем меньше сопротивление R, а величины $\bar{\varphi}_L$ и $\bar{\varphi}_C$ могут во много раз превышать напряжение $\bar{\varphi}$ источника переменного тока.

При последовательном соединении складываются сопротивления. Но «сопротивления» емкости и индуктивности имеют противоположные знаки и поразному зависят от частоты. При резонансной частоте они равны по абсолютной величине и, следовательно, взаимно уничтожаются.

Таким образом, система из последовательно включенных L, R, C имеет при резонансе ($\omega=\omega_0$) минимальное сопротивление, пропускает наибольший ток при данной амплитуде переменного напряжения (по сравнению с током при той же амплитуде напряжений, но при частоте $\omega\neq\omega_0$).

Интересно исследовать подробно, как меняются амплитуды напряжения и силы тока, если мы отходим от точного резонанса, т. е. если $\omega\neq\frac{1}{\sqrt{LC}}$. Для этого поступим следующим образом. Пользуясь формулой (3), найдем, что $\bar{\varphi}=\frac{b}{\sqrt{2}}$, откуда $b=\sqrt{2}\bar{\varphi}$. Подставляя это значение в (5), получим

$$\bar{\varphi}=\frac{j_0}{\sqrt{2}}\sqrt{R^2+\left(L\omega-\frac{1}{C\omega}\right)^2}.$$

Но $\frac{j_0}{\sqrt{2}}=\bar{j}$, поэтому

$$\bar{\varphi}=\bar{j}\sqrt{R^2+\left(L\omega-\frac{1}{C\omega}\right)^2}.$$

Выражая отсюда $\bar{j}$ и подставляя это $\bar{j}$ в (9) и (10), найдем:

$$\bar{\varphi}_L=\frac{L\omega}{\sqrt{R^2+\left(L\omega-\frac{1}{C\omega}\right)^2}}\bar{\varphi},$$

$$\bar{\varphi}_C=\frac{1}{C\omega}\frac{\bar{\varphi}}{\sqrt{R^2+\left(L\omega-\frac{1}{C\omega}\right)^2}}.$$

Обозначим через ω_0 собственную частоту контура. Тогда $\omega_0^2=\frac{1}{LC}$ и последние формулы можно записать так:

$$\bar{\varphi}_L=\frac{\omega^2\bar{\varphi}}{\sqrt{\frac{R^2\omega^2}{L^2}+(\omega_0^2-\omega^2)^2}},$$

$$\bar{\varphi}_C=\frac{\omega_0^2\bar{\varphi}}{\sqrt{\frac{R^2\omega^2}{L^2}+(\omega_0^2-\omega^2)^2}}. \tag{13.13.11}$$

В таком виде ясна зависимость амплитуды от того, насколько близка собственная частота ω_0 контура к частоте ω переменного тока.

Зависимость отношения $\frac{\bar{\varphi}_L}{\bar{\varphi}}$ от ω вблизи точки $\omega=\omega_0$ показана на рис. 2. График построен для случая $\frac{R}{L\omega_0}=0{,}05$. На графике виден типичный так называемый *резонансный ход* кривой.

Если $\frac{R}{L\omega}\ll 1$, то зависимости $\frac{\bar{\varphi}_L}{\bar{\varphi}}$ и $\frac{\bar{\varphi}_C}{\bar{\varphi}}$ от ω определяются главным образом вторым членом подкоренного выражения, т. е. величиной $(\omega_0^2-\omega^2)^2$. При $\omega=\omega_0$ этот член обращается в нуль, при сделанном предположении $\left(\frac{R}{L\omega}\ll 1\right)$ знаме-

натель имеет минимум, а величина амплитуды максимальна. Амплитуда составляет 70% максимальной при $(\omega_0^2-\omega^2)^2=\frac{R^2\omega^2}{L^2}$, т. е. при $\omega_0^2-\omega^2=$ $=\pm\frac{R\omega}{L}$, откуда

$$\omega_0-\omega=\pm\frac{R}{L}\frac{\omega}{\omega_0+\omega}\approx\pm\frac{R}{2L}.$$

То изменение частоты, при котором квадрат амплитуды падает вдвое по сравнению со своим максимальным значением, называется *шириной резонансной кривой* (или *шириной резонанса*). Если амплитуда составляет 70%, т. е. 0,7 от максимальной, то квадрат амплитуды составляет $0{,}7^2\approx 0{,}5$ от максимального. Поэтому ширина резонансной кривой $\omega-\omega_0$ есть $\frac{R}{2L}$, т. е. ширина равна величине, характеризующей скорость затухания колебаний в таком контуре (см. § 9).

Следовательно, чем меньше сопротивление R, тем меньше ширина резонанса, тем круче идет кривая вблизи $\omega=\omega_0$. Из формул (11) видно, что величина максимума $\frac{\varphi_L}{\varphi}$ тем больше, чем меньше R. Поэтому явление резонанса особенно сильно проявляет себя в том случае, когда R мало.

§ 14. Параллельное включение индуктивности и емкости. Резонанс токов

Рассмотрим схему рис. 1, отличающуюся от схемы рис. 13.1 тем, что L и C включены параллельно; сопротивление цепи мы теперь считаем весьма малым и пренебрегаем им. В этом случае φ_C и φ_L одинаковы и равны напряжению в цепи (т. е. в источнике переменного тока) φ, а ток j складывается из тока j_C, текущего через C, и тока j_L, текущего через L. Пусть

$$\varphi_L=\varphi_C=\varphi=\varphi_0\cos(\omega t+\alpha).$$

Пользуясь формулами § 11, найдем:

$$j_C=-C\omega\varphi_0\sin(\omega t+\alpha),$$

$$j_L=\frac{\varphi_0}{L\omega}\sin(\omega t+\alpha).$$

Поэтому

$$j=j_C+j_L=\varphi_0\left(\frac{1}{L\omega}-C\omega\right)\sin(\omega t+\alpha).$$

Отсюда, полагая $\omega_0=\frac{1}{\sqrt{LC}}$, имеем

$$\bar{j}=\bar{\varphi}\left(\frac{1}{L\omega}-C\omega\right),$$

$$\bar{\varphi}=\frac{\bar{j}}{1/L\omega-C\omega}=\frac{\omega\bar{j}}{C(\omega_0^2-\omega^2)}.$$

В этом случае также проявляется типичная резонансная зависимость: при данной силе тока $\bar{j}$ напряжение $\bar{\varphi}$ тем

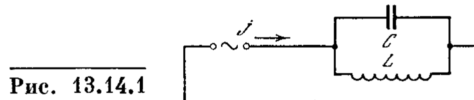


Рис. 13.14.1

больше, чем ближе ω к ω_0. Легко убедиться, что при ω, близком к ω_0, $\bar{j}_L$ и $\bar{j}_C$ гораздо больше, чем j цепи, т. е. в контуре, состоящем из L и C, происходят сильные колебания. При этом достаточно малого внешнего тока для того, чтобы поддерживать гораздо более сильные токи в контуре.

При параллельном соединении, как известно, складываются проводимости, т. е. величины, обратные сопротивлениям:

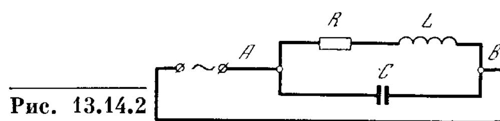


Рис. 13.14.2

$$\frac{1}{R}=\frac{1}{R_1}+\frac{1}{R_2}+\frac{1}{R_3}+\ldots$$

Проводимости (т. е. отношения тока к разности потенциалов) у емкости и индуктивности имеют противоположные знаки и по-разному зависят от частоты. При резонансе ($\omega=\omega_0$) они взаимно уничтожаются и общая проводимость будет наименьшей, т. е. ток будет наименьший при данной разности потенциалов, а, следовательно, разность потенциалов φ_{AB} наибольшая при данном токе во внешней цепи.

В упрощенной схеме без сопротивления амплитуда колебаний неограниченно растет при приближении ω к ω_0. В действительности включенное в схему сопротивление приводит к тому, что амплитуда конечна и при $\omega=\omega_0$.

Если включить R параллельно L и C, то все расчеты становятся очень похожими на расчеты предыдущего параграфа. Но этот случай редко встречается в практике. В действительности

индуктивность обычно имеет заметное сопротивление, поэтому типичной является схема рис. 2. В этом случае расчеты несколько более длинны, чем в предыдущем параграфе, и мы не будем приводить их в подробностях. В результате расчетов (при ω, близком к ω_0, и при малом $\frac{R}{L\omega}$) получим

$$\frac{\bar{j}_L}{\bar{j}} \approx \frac{\bar{j}_C}{\bar{j}} = \frac{\omega_0^2}{\sqrt{(R\omega/L)^2 + (\omega_0^2 - \omega^2)^2}}.$$

Таким образом, оказывается, что усиление тока при резонансе в цепи с параллельным включением подчиняется тому же закону, что и усиление напряжения при последовательном включении, рассмотренное в предыдущем параграфе.

§ 15. Общие свойства резонанса линейной системы

Подчеркнем, что в наших рассуждениях, складывая колебания, вызванные воздействием силы в различные моменты времени, мы все время использовали тот факт, что сама система без внешнего воздействия является л и н е й н о й. Ее поведение описывается линейным дифференциальным уравнением, искомая величина во все члены уравнения входит в первой степени. Поэтому имеет место принцип суперпозиции — сумма двух или нескольких частных решений уравнения также является решением.

В § 13 мы получили формулу для ширины резонансной кривой $\omega - \omega_0 = \frac{R}{2L}$, которая показывает, что чем медленнее затухают колебания, тем меньше ширина резонанса. Этот факт имеет место не только для электрических колебаний. Рассмотрим любую систему, в которой могут происходить колебания. Пусть некоторая внешняя сила вызвала колебания в такой системе, а затем перестала действовать. Система оказалась предоставленной самой себе. После этого колебания начинают затухать. При этом если амплитуда колебаний убывает как $e^{-\gamma t}$, то величина γ характеризует скорость затухания колебаний. Эта величина имеет размерность с$^{-1}$. За время $\tau = \frac{1}{\gamma}$ амплитуда убывает в e раз, т. е. на 63%.

Теперь рассмотрим раскачку такой системы периодической внешней силой. Амплитуда колебаний в данный момент есть сумма амплитуд, приобретенных за все время раскачки. При наличии затухания амплитуда, приобретенная давно, слишком рано успевает затухнуть, не играет роли, не вносит заметного вклада в амплитуду колебаний в данный момент.

Время затухания, очевидно, равно $\frac{1}{\gamma}$, за это время амплитуда свободных колебаний успевает уменьшиться в e раз, т. е. на 63%. Следовательно, даже в том случае, когда раскачивающая сила действует непрерывно, от $t = -\infty$, все равно амплитуда колебаний определяется лишь интервалом времени от $t - \frac{1}{\gamma}$ до t, где t — момент наблюдения; действие силы в более ранний период успевает заглохнуть.

Для того чтобы сильно проявилось различие между двумя периодическими силами с несколько различными периодами, $F_0 \sin \omega_0 t$ и $F_0 \sin \omega t$, нужно такое время наблюдения T, за которое их фазы разойдутся приблизительно на π единиц: $\omega T = \omega_0 T \pm \pi$, так что $|\omega - \omega_0| = \frac{\pi}{T}$. Следовательно, если колебания системы «помнят» только действие силы за время $T = \frac{1}{\gamma}$, то у такой системы различие в частотах возбуждающей силы, меньшее, чем π/T (случай $|\omega - \omega_0| < \frac{\pi}{T} = \pi\gamma$), мало повлияет на амплитуду. Отсюда видно, что ширина резонансной кривой пропорциональна затуханию γ.

С другой стороны, так как система «помнит» и собирает воздействие силы за время $\frac{1}{\gamma}$, то сама амплитуда в резонансе, а значит, и высота резонансной кривой обратно пропорциональны γ. Расчеты подтверждают эти соображения.

Еще раз напомним, что мы все время используем то обстоятельство, что сама система без внешнего воздействия является л и н е й н о й. Поведение такой системы описывается линейным дифференциальным уравнением, в силу чего здесь и применим сформулированный выше п р и н ц и п с у п е р п о-

зиции. Для нелинейных систем принцип суперпозиции не имеет места, и их теория является гораздо более сложной. Укажем еще, что линейная система не генерирует колебаний даже при наличии в схеме ЭДС — (идеальных) источников постоянного (во времени) напряжения.

§ 16*. Ток смещения и электромагнитная теория света

Выше везде мы почти без оговорок рассматривали ток, идущий через конденсатор. Действительно, если включить конденсатор в цепь переменного тока последовательно с амперметром, то амперметр покажет вполне определенную силу тока $\bar{j}=C\omega\bar{\varphi}$. С другой стороны, через конденсатор ток не идет: пластины конденсатора разделены изолятором (например, между ними пустота), поэтому те индивидуальные носители электричества — электроны, которые находятся в левом проводе и на левой пластине, никогда не попадут на правую пластину и правый провод. Следовательно, в пространстве между пластинами не движутся заряженные частицы, т. е. нет электрического тока в том смысле, как мы это понимали до сих пор. В этом пространстве есть только электрическое поле, которое меняется, когда меняется заряд пластин, т. е. когда в левом и правом проводниках течет ток. Теперь возникают две возможности:

1) либо надо извиниться перед читателем, сказав, что везде, где говорилось о токе через емкость, мы выражались не точно — на самом деле тока нет, ток течет только в проводах слева и справа;

2) либо надо рассматривать переменное электрическое поле, возникающее в пространстве между пластинами, наравне с обычным током — движением заряженных частиц.

Максвелл, ставший на эту последнюю точку зрения, сумел получить из нее грандиозные следствия.

Давно было известно, что *электрический ток*, т. е. движение заряженных частиц, вызывает появление *магнитного поля*. Но если переменное электрическое поле подобно электрическому току, то переменное электрическое поле в пустоте тоже должно создавать магнитное поле. Эта гипотеза Максвелла привела к замечательной симметрии между электрическим и магнитным полями. Фарадей экспериментально открыл *индукцию*, т. е. тот факт, что изменение магнитного поля вызывает появление электрического поля. Максвелл умозрительно пришел к предположению о существовании аналогичного явления, при котором изменение электрического поля вызывает появление магнитного поля. Только после этого предположения теория электрических и магнитных полей приобрела современный вид.

Математическая теория Максвелла записывается в виде дифференциальных уравнений, которые слишком сложны для нашей книги, и поэтому мы их приводить не будем [1]. Решения этих уравнений описывают распространение в пустоте электрических и магнитных полей. При этом обязательно должны существовать оба поля: изменение электрического поля вызывает магнитное поле, изменение магнитного поля порождает электрическое поле.

Ко времени, когда работал Максвелл, уже были сделаны опыты Фарадея, известна была связь между переменным магнитым полем и индуцируемой им ЭДС. Известно было магнитное поле тока. Наконец, известна была связь между зарядом конденсатора и электрическим полем между его пластинами. Этих данных достаточно для того, чтобы написать уравнения для полей в пустоте.

Максвелл нашел скорость распространения полей в пустоте. Эта скорость оказалась равной скорости света! Отсюда, естественно, следовало предположение, что свет есть не что иное, как электромагнитные колебания. Далее теория предсказывала возможность

[1] Характеристики электрического и магнитного полей в пространстве являются функциями нескольких переменных — они зависят от трех координат точки в пространстве и от времени. Соответственно этому фигурирующие в уравнениях Максвелла производные суть частные производные функций нескольких переменных, а сами эти уравнения являются так называемыми *дифференциальными уравнениями в частных производных* (ср. со сказанным по этому поводу в Заключении). К тому же эти характеристики полей являются векторными величинами. Укажем еще, что изящные применения в теории электромагнетизма получают комплексные числа и аналитические функции комплексной переменной (см. гл. 17).

существования электромагнитных колебаний любой длины волны, в том числе и рентгеновских лучей (длина которых в тысячи раз меньше длины световой волны), и радиоволн с большой длиной волны. Так, работы Фарадея и Максвелла начали тот путь, который завершился открытием радиоволн Герцем и изобретением радио как средства связи нашим соотечественником А. С. Поповым и итальянцем Маркони, а также открытием Рентгена и созданием полной теории электромагнитного поля и его взаимодействия с веществом.

§ 17*. Нелинейное сопротивление и туннельный диод

Рассмотрим двухполюсник — «ящик», похожий на сопротивление в том отношении, что сила проходящего через него тока зависит только от мгновенного значения разности потенциалов. В этом смысле «ящик» не похож ни на индуктивность, где φ зависит от производной $\frac{dj}{dt}$, ни на емкость, где φ за-

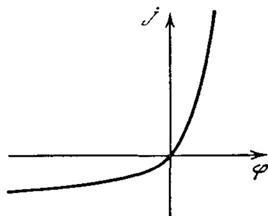


Рис. 13.17.1

висит от интеграла $\int j dt$. Однако и от обычного сопротивления «ящик» отличается тем, что зависимость $j(\varphi)$ не совпадает с элементарным законом Ома $j=\frac{\varphi}{R}$: у «ящика» зависимость $j(\varphi)$ является более сложной. Эта функция $j(\varphi)$ называется *характеристикой* «ящика».

Единственное общее утверждение, которое можно сделать о характеристике $j(\varphi)$, это то, что φ и j не могут быть противоположны по знаку, если внутри «ящика» нет батарей или других источников энергии. Если φ и j одного знака, то при прохождении тока энергия внутри «ящика» поглощается, «ящик» отбирает электрическую энергию от той цепи, в которую он включен; в «ящике» эта энергия превращается в тепловую и рассеивается. Так как φ непрерывно и $\varphi<0$ при $j<0$, $\varphi>0$ при $j>0$, то $\varphi=0$ при $j=0$. В остальном зависимость $j(\varphi)$ может быть любой. Например, в качестве выпрямителей тока применяются «ящики», характеристика которых изображена на рис. 1: в одну

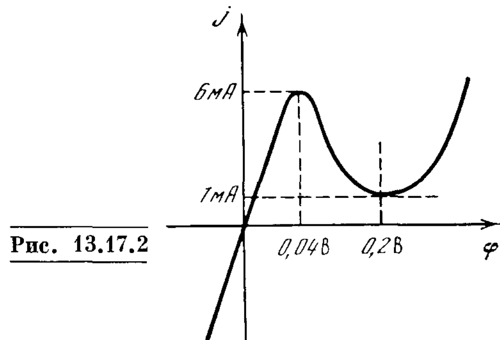


Рис. 13.17.2

сторону ток течет легко при малой разности потенциалов; в другую сторону ток почти не идет — как видно из графика, даже при большой отрицательной разности потенциалов ток мал. Та-

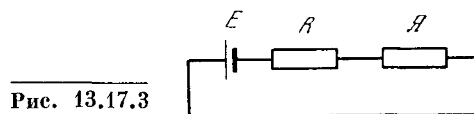


Рис. 13.17.3

кими свойствами обладают так называемые *диоды*, изготовляемые из двух полупроводников. В 1958 г. в Японии удалось сделать из специально подобранных полупроводников «ящик», так

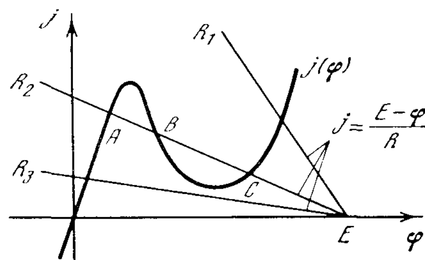


Рис. 13.17.4

называемый *туннельный диод* (в действительности этот «ящик» представляет собой миниатюрный цилиндрик диаметром и высотой всего в несколько миллиметров [1]), с необычной кривой

[1] В конце 70-х годов и физики, и конструкторы очень успешно продвинулись в направлении миниатюризации всех частей электронной аппаратуры. Современный диод (или емкость, или сопротивление) может быть изготовлен в виде пятнышка диаметром в доли миллиметра, состоящего из слоев вещества с нужными свойствами, толщиной, равной сотым или тысячным долям миллиметра.

$j(\varphi)$, имеющей минимум (см. рис. 2, на котором даны типичные значения φ, j). Такая кривая не противоречит высказанному принципу, знак φ везде тот же, что и знак j, так что «ящик» только поглощает энергию. Мы не будем входить в рассмотрение физических причин такой странной кривой, но зато рассмотрим следствия для цепи, в которую включен туннельный диод. Для краткости будем называть его и дальше просто «ящик».

Начнем с наиболее простой цепи, состоящей из трех звеньев: батарея с ЭДС E—сопротивление R (обычное, подчиняющееся закону Ома)—«ящик» $Я$ (рис. 3).

Внутреннее сопротивление батареи включим в R.

Уравнение, определяющее ток и распределение потенциала в цепи, имеет вид

$$-E+Rj(\varphi)+\varphi=0,$$

где φ — разность потенциалов на «ящике», $j(\varphi)$ есть функция, определяемая свойствами «ящика» (см. рис. 2). Ток через R равен току через «ящик», поэтому $\varphi_R=Rj=Rj(\varphi)$.

Это уравнение удобно решить графически. Запишем его так:

$$\varphi=E-Rj(\varphi) \qquad (13.17.1)$$

и построим в плоскости φ, j прямую $\varphi=E-Rj$. Эту прямую можно назвать «нагрузочной линией» системы батарея —сопротивление. Решение задачи дается пересечением прямой (1) (т. е. прямой $j=\frac{E-\varphi}{R}$) на плоскости с координатами φ, j с линией $j(\varphi)$ — характеристикой «ящика». На рис. 4 показано такое графическое решение задачи, притом для одной батареи и трех разных сопротивлений: малого R_1, среднего R_2 и большого R_3.

Из графика видно, что при достаточно большом E можно подобрать такое R — не слишком малое и не слишком большое, — при котором есть три точки пересечения A, B, C, т. е. три решения уравнения (1).

Необходимое условие для существования трех решений — наличие падающего участка кривой $j=j(\varphi)$. Ясно, что линия, вдоль которой всюду $\frac{dj}{d\varphi}>0$, может только один раз пересечься с нагрузочной прямой, каковы бы ни были E и $R>0$.

Рассмотрим теперь несколько усложненную схему с емкостью, включенной параллельно «ящику» (рис. 5). Для такой схемы получим, что ток через «ящик» $j(\varphi)$ и ток через емкость $C\frac{d\varphi}{dt}$

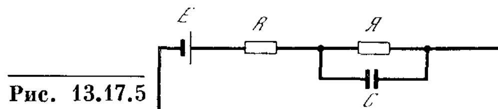


Рис. 13.17.5

в сумме равны току через сопротивление:

$$j(\varphi)+C\frac{d\varphi}{dt}=\frac{E-\varphi}{R}.$$

Отсюда

$$\frac{d\varphi}{dt}=\frac{1}{C}\left[\frac{E-\varphi}{R}-j(\varphi)\right].$$

Точки пересечения характеристики $j=j(\varphi)$ «ящика» с нагрузочной линией $j=\frac{E-\varphi}{R}$ соответствуют решениям $\varphi=\mathrm{const}$, $\frac{d\varphi}{dt}=0$. Выясним, каков знак

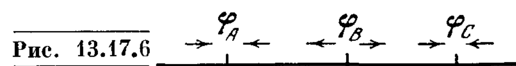


Рис. 13.17.6

$\frac{d\varphi}{dt}$ в окрестности этих точек. Глядя на рис. 4, легко убедиться, что $\frac{d\varphi}{dt}>0$ при $\varphi<\varphi_A$; $\frac{d\varphi}{dt}>0$ при $\varphi_B<\varphi<\varphi_C$; $\frac{d\varphi}{dt}<0$ при $\varphi_A<\varphi<\varphi_B$; $\frac{d\varphi}{dt}<0$ при $\varphi>\varphi_C$.

На рис. 6 стрелками схематически показано направление изменения φ при его изменении в окрестности стационарных значений φ_A, φ_B и φ_C. Из него видно, что среднее решение B неустойчиво: стоит от него немного отойти влево или вправо и появляется $\frac{d\varphi}{dt}$ такого знака, что отклонение φ от φ_B будет увеличиваться! Напротив, точки A и C рис. 4 характеризуют два устойчивых решения, которым соответствуют устойчивые состояния системы.

Наличие двух устойчивых состояний позволяет применять такие «ящики» в математических машинах как ячейки памяти: сделав много таких цепей и переводя внешним воздействием одни

ячейки в состояние A, другие — в состояние C, мы можем «записать», «запомнить» какое угодно число или другую информацию. С помощью таких систем любую информацию мы записываем шифром вида $AACACACCC\ldots$, где каждая буква A или C обозначает состояние, в котором находится соответствующая система (1-я и 2-я — в A, 3-я — в C, 4-я — в A и т. д.).

Рассмотрим теперь систему (рис. 7), состоящую из индуктивности и емкости, включенных параллельно «ящику». Снова обозначаем φ разность потенциалов на «ящике», $j=j(\varphi)$ — его характе-

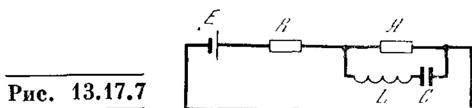


Рис. 13.17.7

ристику; ток, текущий через L и C, обозначим через j_1:

$$j(\varphi)+j_1=\frac{E-\varphi}{R}, \quad C\frac{d\varphi_C}{dt}=j_1,$$

$$\varphi-\varphi_C=\varphi_L=L\frac{dj_1}{dt}. \qquad (13.17.2)$$

Рассмотрим процесс в контуре, когда ток и потенциал в «ящике» близки к средней точке пересечения B. Пусть

$$\varphi=\varphi_B+f, \quad \varphi_C=\varphi_B+g,$$

$$j\approx j(\varphi_B)+f\left.\frac{dj}{d\varphi}\right|_{\varphi=\varphi_B}=j(\varphi_B)+kf,$$

где k — обозначение производной $\frac{dj}{d\varphi}$ при $\varphi=\varphi_B$. Силу тока мы представили первыми двумя членами ряда Тейлора, средний потенциал емкости равен φ_B, причем φ_B и $j(\varphi_B)$ удовлетворяют условию $j(\varphi_B)=\frac{E-\varphi_B}{R}$. Подставляя эти выражения в уравнения (2), получим после сокращений:

$$kf+j_1=-\frac{f}{R}, \quad C\frac{dg}{dt}=j_1,$$

$$f-g=L\frac{dj_1}{dt}. \qquad (13.17.3)$$

Из (3) следует

$$f=-\frac{1}{1/R+k}j_1=-rj_1, \qquad (13.17.4)$$

где $\frac{1}{r}=\frac{1}{R}+k=\frac{1}{R}+\left.\frac{dj}{d\varphi}\right|_{\varphi=\varphi_B}$. Из уравнений (3) и (4) найдем также

$$L\frac{d^2j_1}{dt^2}=\frac{df}{dt}-\frac{dg}{dt}=-r\frac{dj_1}{dt}-\frac{1}{C}j_1$$

и окончательно

$$\frac{d^2j_1}{dt^2}+\frac{r}{L}\frac{dj_1}{dt}+\frac{1}{LC}j_1=0. \qquad (13.17.5)$$

Уравнение (5) представляет собой обычное уравнение колебательного контура с емкостью C, индуктивностью L и сопротивлением r. Сопротивление r соответствует тому, что емкость и индуктивность замкнуты параллельно двумя цепями: цепью батареи с сопротивлением R и «ящиком» с сопротивлением $\frac{1}{k}=\frac{d\varphi}{dj}$. Так как эти две цепи включены параллельно, то складываются проводимости, т. е. величины, обратные сопротивлениям, откуда и следует выражение для r.

Токи и потенциалы в системе распадаются на сумму двух слагаемых: постоянного $(\varphi_B,\ j(\varphi_B))$ и колебательного $(j_1(t),\ f(t),\ g(t))$. При этом для колебательного слагаемого роль сопротивления «ящика» играет производная $\frac{d\varphi}{dj}$, взятая по характеристике «ящика». Будь в «ящике» обычное омическое сопротивление, то мы имели бы соотношение $\varphi=Rj$, т. е. производная и равнялась бы величине сопротивления:

$$\frac{d\varphi}{dj}=R.$$

К чему приводит необычная характеристика $j(\varphi)$ «ящика» туннельного диода (см. рис. 4)? В точке B производная $\frac{dj}{d\varphi}<0$, т. е. по отношению к колебаниям «ящик» имеет отрицательное сопротивление! Более того, из рис. 4 видно, что в точке пересечения $|k|=\left|\frac{dj}{d\varphi}\right|>\frac{1}{R}$, так как $\frac{1}{R}$ — это как раз угловой коэффициент нагрузочной прямой $j=\frac{E-\varphi}{R}$, пересекающей характеристику в точке B. Следовательно, суммарное сопротивление колебательного контура

$$r=\frac{1}{R}+k<0.$$

Уравнение для колебательного контура с L, C, r при $r>0$ давало затухающие колебания. При $r<0$ это уравнение даст раскачивающиеся, усиливающиеся с течением времени колебания.

Таким образом, с помощью туннельного диода можно генерировать (возбуждать) в контуре колебания.

Способность к возбуждению колебаний является следствием неустойчивости точки B. Энергия колебаний черпается от батареи. Амплитуда колебаний растет со временем по экспоненциальному закону лишь до тех пор, пока можно считать ее малой и пользоваться разложением в ряд Тейлора характеристики $j(\varphi)$ около точки B. Максимальная амплитуда, грубо говоря, ограничивается точками A и C на рис. 4.

Уже к 1961 г. были испытаны генераторы с КПД до 25% при мощности 0,5 МВт при частоте 7500 МГц (длина волны 4 см). Для нас цепи с туннельным диодом интересны с точки зрения математического рассмотрения нелинейной задачи, вопросов мощности и представления токов в системе как наложения постоянного решения и колебаний.

ЧАСТЬ ТРЕТЬЯ

ДОПОЛНИТЕЛЬНЫЕ ТЕМЫ ИЗ ВЫСШЕЙ МАТЕМАТИКИ

Глава 14

КОМПЛЕКСНЫЕ ЧИСЛА

§ 1. Основные свойства комплексных чисел

Дополнительный свет на ряд из затронутых в этой книге вопросов проливает использование так называемых *комплексных чисел*. Первоначально число (натуральное) характеризовало количество предметов в каком-либо (конечном) их наборе: количество детей в семье, лодок на реке, пальцев на руке. Числа можно складывать: если объединить две группы лиц из a и из b человек, то всего мы получим $a+b$ человек; их можно также перемножать: a кучек по b плодов в каждой дадут ab плодов. Но вот вычитать и делить натуральные числа можно н е в с е г д а — ограничиваясь одними лишь натуральными числами, нельзя из 5 вычесть 8 или 3 разделить на 2. Стремление сделать операции вычитания и деления всегда выполнимыми привело к появлению *отрицательных* и *дробных* (*рациональных*) *чисел*: условились обозначать разность $5-8$ через -3 и частное $3:2$ через $\frac{3}{2}$. Новые «числа» имели уже совсем другой смысл, чем раньше, ибо ни в одной группе людей количество их не может равняться ни -3, ни $\frac{3}{2}$ (недаром С. Маршак с шуточным ужасом писал о школьнике, у которого в задаче «получилося в ответе два землекопа и две трети»). Также и действия над этими числами определяются уже по-новому: в самом деле, ведь бессмысленно говорить, скажем, просто об объединении в одну флотилию -3 лодок и $\frac{3}{2}$ лодок. Так, например, приходится считать, что сумма дробей $\frac{a}{b}$ и $\frac{c}{d}$ равна $\frac{ad+bc}{bd}$, а произведение $(-a)$ на $(-b)$ равно $+ab$[1].

Возможность неограниченного умножения чисел влечет за собой выполнимость операции возведения числа в (целую положительную) степень: $a^2=a\cdot a$, $a^3=a\cdot a\cdot a$, и вообще $a^n=\underbrace{a\cdot a\cdot a\ldots a}_{n\text{ раз}}$. Но вот обратная операция извлечения корня в области «обычных» чисел выполнима не всегда: ведь не существует никакого (положительного или отрицательного!) числа x, квадрат которого был бы равен, скажем, -7, т. е. не существует числа $x=\sqrt{-7}$. В дальнейшем «обычные» числа, т. е. положительные, отрицательные и нуль, мы будем называть *вещественными*[2]. Стремление сделать выполнимой также и операцию извлечения (квадратного) корня

[1] Эти соглашения диктуются тем, что только при принятии их действия над «новыми» числами целиком охватывают ранее известные действия над «старыми» числами (при расширении понятия числа нам приходится доучиваться, но не переучиваться — так, по нашим правилам, $a+b=\frac{a}{1}+\frac{b}{1}=\frac{a\cdot 1+1\cdot b}{1\cdot 1}=\frac{a+b}{1}=a+b$), а также сохраняют силу все общие правила действий (вроде законов $a+b=b+a$ или $(a+b)\,c=ac+bc$). Те же соображения играют основную роль и при определении правил действий с *комплексными* числами, о которых говорится в настоящей главе.

[2] Для того чтобы сделать возможным извлечение корня квадратного из любого положительного числа, приходится включать в множество вещественных чисел не только все (положительные и неположительные) дроби (рациональные числа), но также и так называемые *иррациональные* числа, вроде числа $\sqrt{2}=1{,}41\ldots$ (По этому поводу см., например, рассчитанную на широкий круг читателей книгу: *Нивен А*. Числа рациональные и иррациональные. М.: Мир, 1966.)

из любого вещественного числа приводит к понятию комплексных чисел.

Введем в рассмотрение новое «число» $\sqrt{-1}$, определяемое тем условием, что его квадрат равен числу -1. Ясно, что ни одно вещественное число — ни положительное, ни отрицательное — такого квадрата не имеет; поэтому $\sqrt{-1}$ мы назовем *мнимой единицей*; в математике ее обычно обозначают буквой i [3]. Выражения вида $c=a+ib$ или $z=x+iy$, где a и b (или x и y) — вещественны, и называют *комплексными числами*; числа вида ib $(=0+ib)$ или iy иногда называют *чисто мнимыми* (или просто *мнимыми*).

Четыре арифметические действия с комплексными числами $a=a_1+ia_2$ и $b=b_1+ib_2$ непосредственно вытекают из определения мнимой единицы. Они вводятся как естественное обобщение обычных действий с вещественными числами: если $a+b=f$, $a-b=g$ и $ab=h$, то

$$f=f_1+if_2=(a_1+ia_2)+(b_1+ib_2)=$$
$$=(a_1+b_1)+i(a_2+b_2);$$
$$g=g_1+ig_2=(a_1+ia_2)-(b_1+ib_2)=$$
$$=(a_1-b_1)+i(a_2-b_2);$$
$$h=h_1+ih_2=(a_1+ia_2)(b_1+ib_2)=$$
$$=a_1b_1+ia_1b_2+ia_2b_1+ia_2ib_2=$$
$$=(a_1b_1-a_2b_2)+i(a_1b_2+a_2b_1)$$

(при вычислении h мы учитываем, что $i^2=-1$); поэтому

$$f_1=a_1+b_1,\quad f_2=a_2+b_2;$$
$$g_1=a_1-b_1,\quad g_2=a_2-b_2;$$
$$h_1=a_1b_1-a_2b_2,\quad h_2=a_1b_2+a_2b_1.$$

Аналогично, если $k=\frac{a}{b}=k_1+ik_2$, то $a=bk$, т. е.

$$(b_1+ib_2)(k_1+ik_2)=$$
$$=(b_1k_1-b_2k_2)+i(b_1k_2+b_2k_1)=a_1+ia_2.$$

Теперь для определения k_1 и k_2 мы имеем два линейных уравнения:

$$b_1k_1-b_2k_2=a_1,\quad b_2k_1+b_1k_2=a_2,$$

которые можно решить во всех случаях, когда $b\neq 0$ (т. е. когда b_1 и b_2 не равны нулю одновременно):

$$k_1=\frac{a_1b_1+a_2b_2}{b_1^2+b_2^2},\quad k_2=\frac{a_2b_1-a_1b_2}{b_1^2+b_2^2}.$$

[3] От французского слова imaginaire — мнимый, воображаемый, несуществующий.

При нахождении частного $k=\frac{a}{b}$ можно также воспользоваться тем, что при любом (комплексном) числе $b=b_1+ib_2$ произведение b и числа $b^*=b_1-ib_2$ (число b^* часто обозначают через $\bar{b}$; его называют *сопряженным* с b) всегда в е щ е с т в е н н о:

$$bb^*=(b_1+ib_2)(b_1-ib_2)=$$
$$=b_1^2-(ib_2)^2=b_1^2-i^2b_2^2=b_1^2+b_2^2. \quad \textbf{(14.1.1)}$$

(Если $b^*=b$, то число b вещественно; корень квадратный из произведения $\sqrt{bb^*}=\sqrt{b_1^2+b_2^2}$ обозначается через $|b|$ и называется *абсолютной величиной*, или *модулем*, комплексного числа b.) Поэтому

$$k=\frac{a}{b}=\frac{a_1+ia_2}{b_1+ib_2}=\frac{(a_1+ia_2)(b_1-ib_2)}{(b_1+ib_2)(b_1-ib_2)}=$$
$$=\frac{(a_1b_1+a_2b_2)+i(a_2b_1-a_1b_2)}{b_1^2+b_2^2}=$$
$$=\frac{a_1b_1+a_2b_2}{|b|^2}+i\frac{a_2b_1-a_1b_2}{|b|^2}.$$

Наши правила позволяют представить любую рациональную функцию комплексной переменной $z=x+iy$, например $f=\frac{z}{1+z^2}$ и т. п., в виде суммы $f=u+iv$, где u и v — две вещественные функции двух вещественных переменных x и y (см. упр. 1).

Аналогично решается задача и об извлечении корня квадратного из комплексного числа. В самом деле,

если $r=r_1+ir_2=\sqrt{a}=\sqrt{a_1+ia_2}$,

то $(r_1+ir_2)^2=(r_1^2-r_2^2)+i2r_1r_2=a=$
$=a_1+ia_2$,

что приводит к следующей системе уравнений для определения неизвестных r_1, r_2:

$$r_1^2-r_2^2=a_1,\quad 2r_1r_2=a_2.$$

Эта система легко решается: поскольку $r_1^2+(-r_2^2)=a_1$ и $r_1^2(-r_2^2)=-\frac{a_2^2}{4}$, то r_1^2 и $-r_2^2$ являются корнями квадратного уравнения $X^2-a_1X-\frac{a_2^2}{4}=0$, откуда

$$r_1=\sqrt{\frac{|a|+a_1}{2}},\quad r_2=\sqrt{\frac{|a|-a_1}{2}},$$

где, как всегда, $|a|=\sqrt{a_1^2+a_2^2}$. (Заметьте, что оба подкоренных выражения положительны!) Итак,

$$\sqrt{a}=\sqrt{a_1+ia_2}=$$
$$=\pm\left(\sqrt{\frac{|a|+a_1}{2}}+i\sqrt{\frac{|a|-a_1}{2}}\right).$$

Таким образом, вводя новое, «мнимое» число i с тем, чтобы задача извлечения корня квадратного из -1 оказалась разрешимой, мы неожиданно получили возможность извлекать квадратные корни сразу из всех чисел — «старых», вещественных, и «новых», комплексных.

Прежде чем идти дальше, укажем геометрическое представление комплексных чисел. Комплексное число $z=x+iy$ принято рассматривать как точку так называемой *комплексной плоскости* (или *плоскости комплексной переменной*) с (декартовыми прямоугольными) координатами x и y (рис. 1); ось x при этом называют *вещественной осью* (ее точки отвечают вещественным числам $x+i0=x$), а ось y — *мнимой осью*. Каждому комплексному числу отвечает единственная точка плоскости и наоборот; сопряженным комплексным числам отвечают точки, симметричные относительно вещественной оси. Модуль $\sqrt{zz^*}=|z|=\sqrt{x^2+y^2}$ комплексного числа $z=x+iy$ равен расстоянию Oz точки z от начала координат (отвечающего числу 0).

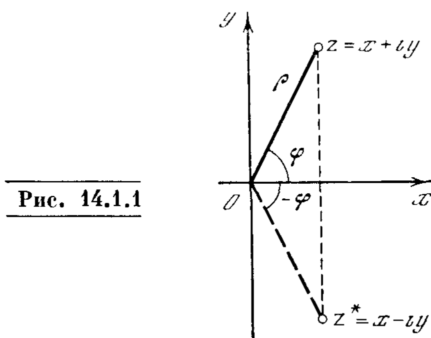


Рис. 14.1.1

Прямую $b_1x+b_2y+c=0$ плоскости комплексной переменной $z=x+iy$ можно задать «комплексным уравнением»

$$bz+b^*z^*+c=0, \tag{14.1.2}$$

где c — вещественно, т. е.

$c^*=c$, а $b={}^1/_2(b_1-ib_2)$.

В самом деле, если $z=x+iy$, то левая часть (2), как нетрудно видеть, равна b_1x+b_2y+c. Уравнение $(x^2+y^2)+b_1x+b_2y+c=0$ *окружности* плоскости комплексной переменной можно переписать в форме, аналогичной (2):

$zz^*+bz+b^*z^*+c=0,$

$$\text{где } c^*=c,\ b={}^1/_2(b_1-ib_2), \tag{14.1.2а}$$

— ведь $zz^*=x^2+y^2$, а $bz+b^*z^*=b_1x+b_2y$.

Уравнения (2) и (2а) можно объединить в одной записи:

$$azz^*+bz+b^*z^*+c=0, \tag{14.1.2б}$$

где a и c — вещественны ($a^*=a$, $c^*=c$); равенство (2б) описывает окружность, если $a\neq 0$, и прямую, если $a=0$.

Положение точки z на комплексной плоскости можно также охарактеризовать ее *полярными координатами* — расстоянием ρ от точки z до начала координат O (это есть модуль $|z|$ числа z) и углом φ, который луч Oz образует с положительным лучом вещественной оси (угол φ называется *аргументом* числа z; иногда его обозначают символом $\operatorname{Arg} z$). При этом (см. рис. 1)

$$x=\rho\cos\varphi,\quad y=\rho\sin\varphi,\quad \rho=\sqrt{x^2+y^2}=$$
$$=|z|,$$
$$\operatorname{tg}\varphi=\frac{y}{x},\quad \cos\varphi=\frac{x}{|z|},\quad \sin\varphi=\frac{y}{|z|}.$$

Таким образом,

$$z=\rho(\cos\varphi+i\sin\varphi) \tag{14.1.3}$$

(*тригонометрическая форма* комплексного числа).

Форма (3) особенно удобна при умножении и делении комплексных чисел

$z_1=\rho_1(\cos\varphi_1+i\sin\varphi_1)$

и $z_2=\rho_2(\cos\varphi_2+i\sin\varphi_2)$.

В самом деле, очевидно,

$$z_1z_2=\rho_1(\cos\varphi_1+i\sin\varphi_1)\rho_2(\cos\varphi_2+$$
$$+i\sin\varphi_2)=\rho_1\rho_2[(\cos\varphi_1+i\sin\varphi_1)\times$$
$$\times(\cos\varphi_2+i\sin\varphi_2)]=\rho_1\rho_2[(\cos\varphi_1\cos\varphi_2-$$
$$-\sin\varphi_1\sin\varphi_2)+i(\cos\varphi_1\sin\varphi_2+$$
$$+\sin\varphi_1\cos\varphi_2)]=\rho_1\rho_2[\cos(\varphi_1+\varphi_2)+$$
$$+i\sin(\varphi_1+\varphi_2)], \tag{14.1.4}$$

или словами: *при умножении комплексных чисел их модули* (абсолютные величины) *перемножаются, а аргументы складываются*. А отсюда в свою очередь вытекает следующее. Если $\frac{z_1}{z_2}=z_3=$ $=\rho_3(\cos\varphi_3+i\sin\varphi_3)$, то $z_1=z_2z_3$,

т. е. $\rho_1=\rho_2\rho_3$, $\varphi_1=\varphi_2+\varphi_3$, откуда $\rho_3=\frac{\rho_1}{\rho_2}$, $\varphi_3=\varphi_1-\varphi_2$. Таким образом,

$$\frac{z_1}{z_2}=\frac{\rho_1}{\rho_2}[\cos(\varphi_1-\varphi_2)+i\sin(\varphi_1-\varphi_2)] \quad (14.1.4а)$$

— *при делении комплексных чисел их модули делятся, а аргументы вычитаются.*

Заметим, что правила

$$|z_1z_2|=|z_1||z_2|, \quad \left|\frac{z_1}{z_2}\right|=\frac{|z_1|}{|z_2|} \quad (14.1.5)$$

в точности совпадают с теми, которые имеют место для вещественных чисел. А вот правила, относящиеся к аргументам комплексных чисел:

$$\operatorname{Arg}(z_1z_2)=\operatorname{Arg}z_1+\operatorname{Arg}z_2,$$
$$\operatorname{Arg}\left(\frac{z_1}{z_2}\right)=\operatorname{Arg}z_1-\operatorname{Arg}z_2, \quad (14.1.5а)$$

являются для нас новыми. По форме равенства (5а) напоминают правила действий с логарифмами, но пока совсем не ясно, какое отношение может иметь угол $\operatorname{Arg}z=\varphi$ к логарифмам (разве у комплексных чисел существуют логарифмы? — ответ на этот вопрос будет дан ниже).

Умножая по формуле (4) число $z=\rho(\cos\varphi+i\sin\varphi)$ само на себя n раз, получим так называемую формулу Муавра:

$$z^n=\rho^n(\cos n\varphi+i\sin n\varphi). \quad (14.1.6)$$

Эта формула справедлива также для отрицательных и для дробных n: так, например, учитывая, что $1=1(\cos 0+i\sin 0)$, получим

$$z^{-1}=\frac{1}{z}=\frac{1}{\rho}[\cos(-\varphi)+i\sin(-\varphi)]=$$
$$=\frac{1}{\rho}(\cos\varphi-i\sin\varphi) \quad (14.1.6а)$$

и

$$z^{1/3}=\sqrt[3]{r(\cos\alpha+i\sin\alpha)}=$$
$$=\sqrt[3]{r}\left(\cos\frac{\alpha}{3}+i\sin\frac{\alpha}{3}\right) \quad (14.1.6б)$$

(см. упр. 2). Заметим, что формула (6б) дает три значения корня кубического из числа z: так как вместо α аргумент z можно также записать в виде $\alpha+2\pi$ или $\alpha+4\pi$, то имеем

$$z_1=\sqrt[3]{r}(\cos\beta+i\sin\beta),$$
$$z_2=\sqrt[3]{r}\left[\cos\left(\beta+\frac{2\pi}{3}\right)+i\sin\left(\beta+\frac{2\pi}{3}\right)\right],$$
$$z_3=\sqrt[3]{r}\left[\cos\left(\beta+\frac{4\pi}{3}\right)+i\sin\left(\beta+\frac{4\pi}{3}\right)\right],$$

где $\beta=\frac{\alpha}{3}$ (рис. 2).

Таким образом, использование комплексных чисел позволяет найти три корня z_1, z_2 и z_3 кубического уравнения $z^3-c=0$, где $c=r(\cos\alpha+i\sin\alpha)$ — произвольное комплексное число. Более того, можно показать, что каждое *алгебраическое уравнение n-й степени*

$$P(z)=a_0z^n+a_1z^{n-1}+a_2z^{n-2}+\ldots$$
$$\ldots+a_{n-1}z+a_n=0,$$

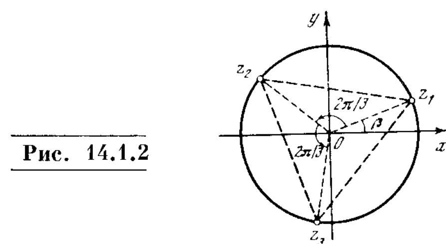


Рис. 14.1.2

где $a_0\neq 0$, a_1, a_2, ..., a_n — произвольные комплексные числа, *имеет n* (и только n), вообще говоря, *комплексных корней* z_1, z_2, ..., z_n, которые, однако, не обязаны быть все различны (некогда это весьма важное предложение называли основной теоремой алгебры). При этом многочлен $P(z)$ представляет собой произведение n линейных множителей:

$$P(z)=a_0(z-z_1)(z-z_2)\ldots(z-z_n).$$

Так, например, квадратное уравнение

$$Q(z)=z^2+pz+q=0 \quad (14.1.7)$$

имеет два корня:

$$z_{1,2}=-\frac{p}{2}\pm\sqrt{\frac{p^2}{4}-q}, \quad (14.1.7а)$$

причем

$$Q(z)=(z-z_1)(z-z_2).$$

Корни (7а) уравнения (7) при вещественных p, q будут вещественными и различными при $\frac{p^2}{4}-q>0$ (напоминаем, что вещественное число $x=x+i0$ является частным случаем комплексного); одинаковыми $\left(\text{равными } -\frac{p}{2}\right)$ при $\frac{p^2}{4}-q=0$; комплексно-сопряженными $\Big($равными $-\frac{p}{2}\pm i\sqrt{q-\frac{p^2}{4}}$, где теперь

подкоренное выражение положительно) при $\frac{p^2}{4} - q < 0$. (Когда говорят, что при вещественных p, q и при $\frac{p^2}{4} < q$ квадратное уравнение (7) не имеет решений, то это не ошибка, а точка зрения: если ограничиться одними лишь вещественными числами, то придется считать, что в этом случае уравнение (7) корней не имеет.) Уравнение

$$z^5 - 3z^4 + 7z^3 - 13z^2 + 12z - 4 = (z-1)^3(z^2+4) = 0 \qquad (14.1.8)$$

имеет пять корней, из которых, однако, только три различны:

$$z_1 = z_2 = z_3 = 1, \quad z_4 = 2i \text{ и } z_5 = -2i.$$

Если все коэффициенты уравнения (любой степени) $a_0z^n + a_1z^{n-1} + \ldots + a_{n-1}z + a_n = 0$ вещественны, то каждому «существенно комплексному» (т. е. не вещественному) корню $z = x + iy$ уравнения отвечает второй корень $z^* = x - iy$, комплексно-сопряженный с первым. Это вытекает из правил действий над комплексными числами, в силу которых

$$(z_1+z_2)^* = z_1^* + z_2^*, \quad (z_1-z_2)^* = z_1^* - z_2^*,$$
$$(z_1z_2)^* = z_1^*z_2^*,$$
$$\left(\frac{z_1}{z_2}\right)^* = \frac{z_1^*}{z_2^*} \qquad (14.1.9)$$

(проверьте!). А в силу (9), если, например, $z_0 = x_0 + iy_0$ — корень уравнения $Q(x) = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4 = 0$ с вещественными коэффициентами, т. е.

$$Q(z_0) = a_0z_0^4 + a_1z_0^3 + a_2z_0^2 + a_3z_0 + a_4 = 0,$$

то

$$a_0^*(z_0^*)^4 + a_1^*(z_0^*)^3 + a_2^*(z_0^*)^2 + a_3^*(z_0^*) + a_4^* = 0^*.$$

Но ведь $a_0^* = a_0$, $a_1^* = a_1$, ..., $a_4^* = a_4$ и уж, конечно, $0^* = 0$; поэтому

$$Q(z_0^*) = a_0(z_0^*)^4 + a_1(z_0^*)^3 + a_2(z_0^*)^2 + a_3(z_0^*) + a_4 = 0,$$

т. е. $z_0^* = x_0 - iy_0$ также является корнем того же самого уравнения (отличным от z_0 при $y_0 \neq 0$).

Из сказанного в свою очередь следует, что каждый *вещественный многочлен* (т. е. такой, все коэффициенты которого вещественны) *можно разложить на* (*вещественные*) *линейные и квадратичные множители* (см., например, разложение (8)). В самом деле, многочлен $P(z) = a_0z^n + a_1z^{n-1} + \ldots + a_{n-1}z + a_n$ имеет n (вещественных или комплексных) корней z_1, z_2, ..., z_n, обращающих его в нуль: $P(z_1) = P(z_2) = \ldots = P(z_n) = 0$. При этом $P(z)$ может быть разложен в произведение

$$P(z) = a_0(z-z_1)(z-z_2)\ldots(z-z_n).$$

Но если $z_0 = x_0 + iy_0$, где $y_1 \neq 0$, — какой-то комплексный корень многочлена $P(z)$, а $z_0^* = x_0 - iy_0$ — корень, комплексно-сопряженный с z_0, то

$$(z-z_0)(z-z_0^*) = (z - x_0 - iy_0)(z - x_0 + iy_0) = [(z-x_0) - iy_0][(z-x_0) + iy_0] = (z-x_0)^2 - (iy_0)^2 = (z-x_0)^2 + y_0^2 = z^2 - 2x_0z + (x_0^2 + y_0^2).$$

Таким образом, каждой *паре комплексно-сопряженных корней* z_0, z_0^* многочлена $P(z)$ отвечает квадратичный множитель $z^2 - 2x_0z + (x_0^2 + y_0^2)$ его разложения на множители, а *вещественному корню* $z_1 = x_1 + i0 = x_1$, если только такой корень существует, отвечает линейный множитель $z - z_1 = z - x_1$ разложения $P(z)$. Это рассуждение и доказывает требуемое утверждение. Так, например, разложение на множители

$$x^4 + a^4 = x^4 + 2a^2x^2 + a^4 - 2a^2x^2 = (x^2+a^2)^2 - (\sqrt{2}\,ax)^2 = [(x^2+a^2) + \sqrt{2}\,ax][(x^2+a^2) - \sqrt{2}\,ax] = (x^2 + \sqrt{2}\,ax + a^2)(x^2 - \sqrt{2}\,ax + a^2)$$

неопытному человеку может показаться случайным и искусственным (откуда мы знали, что к выражению $x^4 + a^4$ надо было прибавить $2a^2x^2$ и отнять $2a^2x^2$?). Однако искушенный читатель поймет, что это разложение с необходимостью вытекает из сформулированного выше общего факта о разложимости вещественных многочленов и из формулы

$$x = \sqrt[4]{-a^4} = a\sqrt[4]{-1}$$

для корней многочлена $x^4 + a^4$ (см. упр. 4).

Тот факт, что в области комплексных чисел корень квадратный извлекается из любого (положительного или отрицательного вещественного или комплексного) числа и что каждое квадратное уравнение имеет здесь два корня, безусловно, радует нас — он доказывает, что множество комплексных чисел в определенном смысле «проще устроено», чем множество вещественных чисел. Однако на этот результат мы хотя бы могли надеяться — ведь мнимая единица i специально вводилась для того, чтобы из числа -1 можно было извлечь корень (или чтобы квадратное уравнение $x^2 + 1 = 0$ можно было решить). Конечно, уже здесь мы получили больше того, на что мы имели основания рассчитывать (существова-

ние не одного лишь ранее несуществующего квадратного корня $\sqrt{-1}$, а всевозможных квадратных корней из всех комплексных чисел; разрешимость не только специально выбранного уравнения $x^2+1=0$, а всех квадратных уравнений), — но это лишь показывало, насколько удачно было соглашение $\sqrt{-1}=i$. Однако то, что введение комплексных чисел $z=x+iy$ позволяет из всех (и «старых» и «новых»!) чисел извлекать корни любой степени (ср. упр. 3), и то, что оно позволяет решать алгебраические уравнения 3, 4, 5-й и каких угодно других степеней, так что, для того чтобы сделать все уравнения разрешимыми, не приходится вводить никаких новых чисел, а можно обойтись теми же комплексными числами, с помощью которых решаются квадратные уравнения,— это, конечно, сюрприз, и притом весьма приятный. Такое положение дела заставляет думать, что в комплексных числах «что-то есть», что в них заложено больше, чем только лишь учение о «корнях из отрицательных чисел», что эти числа вовсе не случайны и могут принести пользу во многих разделах математики и математизированного естествознания.

И эти надежды полностью сбываются.

Упражнения

14.1.1. Пусть $z=x+iy$, а $f(z)=u+iv$; найдите u и v как функции от x и y, если функция $f(z)$ имеет вид: а) $f=z^3-3z+1$; б) $f=\frac{z}{1+z^2}$; в) $f=z^n$.

14.1.2. Докажите формулы (6а), (6б), а также справедливость формулы Муавра (6) для любого (вещественного) n.

14.1.3. Докажите, что n значений корня степени n из произвольного комплексного числа c (n корней уравнения n-й степени $z^n-c=0$) геометрически изображаются n точками комплексной плоскости, являющимися вершинами правильного n-угольника.

14.1.4. Выпишите все значения $\sqrt[4]{-1}$; с помощью полученных формул разложите на (вещественные) множители многочлен z^4+a^4.

§ 2. Возведение в мнимую степень и число e

Поставим задачу определения *мнимой степени*. Как найти число 10^i? Можно ли умножить 10 само на себя i раз, т. е. $\sqrt{-1}$ раз? На первый взгляд сам этот вопрос кажется бессмысленным. Однако ведь и вопрос о том, что получится, если умножить 10 само на себя -3 раза или 1/2 раза, сначала тоже кажется бессмысленным, ибо буквальное выполнение правила возведения числа в целую положительную степень ($a^n=\underbrace{aa\ldots a}_{n\text{ раз}}$) здесь невозможно, а между тем числа $10^{-3}(=0{,}001)$ и $10^{1/2}(=\sqrt{10}\approx\approx 3{,}16)$ безусловно существуют.

Вспомним логическую последовательность определения отрицательных и дробных степеней чисел — этот пример является вдохновляющим! Первоначально мы определяем лишь возведение в целую положительную степень n:

$$a^n=\underbrace{aaa\ldots a}_{n\text{ раз}}.$$

Из этого определения — по-прежнему для целых положительных показателей — следуют два правила:

$$a^{n+m}=\overbrace{aa\ldots a}^{n+m\text{ раз}}=\overbrace{a\ldots a}^{n\text{ раз}}\cdot\overbrace{a\ldots a}^{m\text{ раз}}=a^na^m, \qquad \textbf{(14.2.1)}$$

$$(a^n)^m=\underbrace{\overbrace{a\ldots a}^{n\text{ раз}}\overbrace{a\ldots a}^{n\text{ раз}}\ldots\overbrace{a\ldots a}^{n\text{ раз}}}_{m\text{ раз}}=a^{nm} \qquad \textbf{(14.2.1а)}$$

(ср. со сказанным на с. 408 об определениях сложения и умножения натуральных чисел). Обобщение заключается в том, что мы требуем выполнения правил (1) и (1а) также для отрицательных и дробных показателей степени.

Тогда из правила (1) получаем

$$a^4a^{-3}=a^{4+(-3)}=a^1=a,$$

и, значит,

$$a^{-3}=\frac{a}{a^4}=\frac{1}{a^3},$$

и аналогично для любого натурального m

$$a^{-m}=\frac{1}{a^m}.$$

Таким же образом в силу (1а)

$$(a^{1/3})^3=a^{\frac{1}{3}\cdot 3}=a^1=a,$$

и поэтому

$$a^{1/3}=\sqrt[3]{a} \quad \text{и вообще} \quad a^{1/q}=\sqrt[q]{a}.$$

Теперь для любой дроби $n=p/q$ имеем

$$a^{p/q}=(a^{1/q})^p=\left(\sqrt[q]{a}\right)^p=\sqrt[q]{a^p}. \qquad (14.2.2)$$

Поскольку для любого вещественного (иррационального) числа можно найти как угодно близкую рациональную дробь, задача о возведении в **любую** вещественную степень полностью решена. Однако те приемы, с помощью которых выше удалось определить отрицательную и дробную степени, сами по себе недостаточны для того, чтобы определить *мнимую* степень a^{ik} (a и k — вещественные числа), — для этого подвига нам придется использовать те

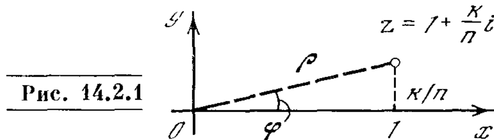


Рис. 14.2.1

сведения, которые мы накопили при изучении производной степенной функции.

Воспользуемся тем, что в силу определения числа e

$$e^r\approx 1+r \quad \text{при} \quad |r|\ll 1 \qquad (14.2.3)$$

(см. § 4.8, формула (4.8.2)), причем приближенное равенство (3) тем точнее, чем меньше $|r|$ [1]. Условимся теперь считать, что (3) сохраняет силу также и для комплексных r, малых по абсолютной величине, в частности и для чисто мнимых чисел ri малого модуля $|ri|=r$. Если мы хотим определить число e^{ki}, где k — какое угодно (не малое!) вещественное число, то воспользуемся тем, что в силу (3)

$$e^{\frac{k}{n}i}\approx 1+\frac{k}{n}i \quad \text{при} \quad n\gg 1, \qquad (14.2.3а)$$

где под n, как это следует из записи $n\gg 1$, понимается **очень большое** число (его мы будем считать **натуральным**), так что $k/n\ll 1$.

Точка $z=1+\frac{k}{n}i$ $(=\rho(\cos\varphi+i\sin\varphi))$ комплексной плоскости имеет декартовы координаты $\left(1,\ \frac{k}{n}\right)$ и полярные координаты $(\rho,\ \varphi)$ (модуль и аргумент комплексного числа z), где

$$\rho=\sqrt{1+\left(\frac{k}{n}\right)^2}=\left(1+\frac{k^2}{n^2}\right)^{1/2},$$

$$\sin\varphi=\frac{k/n}{\rho},\quad \operatorname{tg}\varphi=\frac{k}{n} \qquad (14.2.4)$$

(рис. 1). При $n\gg 1$, т. е. при $\frac{k}{n}\ll 1$ (точные) равенства (4) можно заменить (приближенными) соотношениями

$$\rho\approx 1,\quad \varphi\approx\frac{k}{n}, \qquad (14.2.5)$$

где слова о приближенном характере (5) указывают на пренебрежение слагаемыми порядка малости $\left(\frac{k}{n}\right)^2$ и более высокого (ср. с текстом, напечатанным ниже мелким шрифтом). А так как в силу формулы Муавра (1.6)

$$e^{ki}\approx\left(1+\frac{k}{n}i\right)^n=[\rho(\cos\varphi+i\sin\varphi)]^n=$$
$$=\rho^n(\cos n\varphi+i\sin n\varphi),$$

то

$$\left(1+\frac{k}{n}i\right)^n\approx$$
$$\approx 1^n\left[\cos\left(n\frac{k}{n}\right)+i\sin\left(n\frac{k}{n}\right)\right]=$$
$$=\cos k+i\sin k \qquad (14.2.3б)$$

и, следовательно,

$$e^{ki}=\cos k+i\sin k. \qquad (14.2.6)$$

Это соотношение можно принять за **определение** экспоненциальной функции мнимого аргумента; оно вытекает из предположения (3а). Формула (6) называется **формулой Эйлера**.

Формула Эйлера (6) позволяет ответить на вопрос о том, чему равно число 10^i:

$$10^i=(e^{\ln 10})^i=e^{\ln 10\cdot i}\approx e^{2,3i}=$$
$$=\cos 2,3+i\sin 2,3\approx -0,67+0,77i.$$

[1] Разность $\sigma=e^r-(1+r)$ является малой 2-го порядка по отношению к r: при малых $|r|$ отношение $\frac{\sigma}{r^2}$ остается «конечным», т. е. не очень малым и не очень большим. (Вместо ссылки на равенство (4.8.2) мы могли бы, что равносильно, исходить из определения (4.8.3а) экспоненциальной функции: $e^k=\lim\limits_{n\to\infty}\left(1+\frac{k}{n}\right)^n$, принимая, что последняя формула сохраняет силу и для комплексных (в частности, чисто мнимых) показателей степени k.)

Другой способ вывода формулы Эйлера связан с разложением функций в ряды. Как мы знаем (см. гл. 6),

$$e^x \approx 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots, \tag{14.2.7}$$

причем ряд (7) сходится при всех x, что дает нам смелость считать в (7) число x, если угодно, и комплексным. Положим, по определению,

$$e^{ix} = 1 + \frac{ix}{1!} + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \dots = \left[1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots\right] + i\left[\frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots\right]. \tag{14.2.8}$$

Но поскольку (см. гл. 6)

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots, \tag{14.2.9}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots \tag{14.2.9a}$$

(ряды (9) и (9а) также сходятся при всех x), то мы приходим снова к той же формуле Эйлера:

$$e^{ix} = \cos x + i \sin x.$$

Далее естественно положить:

если $z = x + iy$, то $e^z = e^{x+iy} = e^x e^{iy} = e^x(\cos y + i \sin y)$ (14.2.10)

(ср. упр. 1).

Разумеется, формулу (10) можно сразу принять за определение показательной функции e^z комплексного аргумента $z = x + iy$ (ср. со сказанным о формуле Эйлера (6)). Это определение оправдано тем, что при $z = x + i0 = x$ величина e^z обращается в экспоненциальную функцию e^x вещественной переменной x, а также тем, что для функции (10) сохраняется основное свойство (1) показательной функции:

если $z = z_1 + z_2$, где $z_1 = x_1 + iy_1$, $z_2 = x_2 + iy_2$, то $e^z = \exp[(x_1 + x_2) + i(y_1 + y_2)] = e^{(x_1+x_2)}[\cos(y_1 + y_2) + i\sin(y_1 + y_2)] = (e^{x_1}e^{x_2})[(\cos y_1 + i\sin y_1)(\cos y_2 + i \sin y_2)] = [e^{x_1}(\cos y_1 + i \sin y_1)][e^{x_2}(\cos y_2 + i \sin y_2)] = e^{z_1}e^{z_2}$.

В силу формул (4) и (6.4.6)

$$\rho = \left(1 + \frac{k^2}{n^2}\right)^{1/2} = 1 + \frac{1}{2}\frac{k^2}{n^2} + \dots = 1 + \frac{k^2}{2n^2} + \dots,$$

где точками обозначены члены более высокого порядка малости, чем выписанный член $\frac{k^2}{2n^2}$; в силу же (более общей) формулы (6.4.3)

$$\rho^n = \left(1 + \frac{k^2}{2n^2} + \dots\right)^n = 1 + n\frac{k^2}{2n^2} + \dots = 1 + \frac{k^2}{2n} + \dots,$$

откуда в силу малости при $n \ll 1$ второго слагаемого $\frac{k^2}{2n}$ следует, что модуль числа $\left(1 + \frac{ki}{n}\right)^n$ $(\approx e^{ki})$ вполне допустимо считать (приближенно) равным 1. Аналогично можно проверить, что мы не делаем заметной ошибки и приравняв аргумент числа $\left(1 + \frac{ki}{n}\right)$ величине $\frac{k}{n}$ (ср. упр. 2).

Упражнения

14.2.1. Докажите, что соотношение $e^{x+iy} \approx \left(1 + \frac{x+iy}{n}\right)^n$ при $n \gg 1$ $\left(\text{т.е. } e^{x+iy} = \lim_{n\to\infty}\left(1 + \frac{x+iy}{n}\right)^n\right)$ равносильно формуле (10).

14.2.2. Докажите, что в формуле $1 + \frac{ki}{n} = \rho(\cos\varphi + i \sin\varphi)$ при $n \gg 1$ имеем $\varphi = \frac{k}{n} + \dots$, где точками обозначены члены порядка малости $\frac{1}{n^2}$ и выше.

14.2.3. Докажите, что при любом комплексном (а не только при чисто мнимом) числе Δz, малом по абсолютной величине, $e^{\Delta z} = 1 + \Delta z + \dots$, где точками обозначены члены, модуль которых имеет порядок малости $|\Delta z|^2$ и выше.

§ 3. Тригонометрические функции и логарифмы

Формула Эйлера (2.6) вскрывает глубокую внутреннюю связь периодических функций $\cos x$ и $\sin x$ с показательной функцией. На первый взгляд показательная функция a^x не имеет ничего общего с периодичностью: при $a > 1$ величина a^x монотонно растет с ростом x, а при $a < 1$ она все время убывает; никакой периодичности тут

нет. Но если $a=-1$, то степени a^n числа a принимают значения:

$$-1,\ +1,\ -1,\ +1,\ -1,\ +1,\ \ldots, \quad (14.3.1)$$

т. е. они меняются периодически [1]. Если записать, что

$$-1=\cos\pi+i\sin\pi=e^{i\pi},$$

то мы сможем считать показатель степени n в записи $(-1)^n$ любым вещественным числом:

$$(-1)^n=(e^{i\pi})^n=e^{in\pi}=\cos(n\pi)+i\sin(n\pi).$$

Здесь (вообще говоря, комплексная) величина $(-1)^n$ принимает всевозможные значения, модуль которых равен 1, по-прежнему меняясь периодически с изменением n. Точно так же периодически меняется величина a^x, где $a=\cos\varphi+i\sin\varphi=e^{i\varphi}$ — произвольное комплексное число с единичным модулем: в этом случае

$$a^x=(e^{i\varphi})^x=e^{i\varphi x}=\cos(\varphi x)+i\sin(\varphi x).$$

Если же $a=\rho(\cos\varphi+i\sin\varphi)=\rho e^{i\varphi}$ — комплексное число, модуль которого ρ отличен от 1 ($\rho>0$ и $\rho\neq 1$), то величина

$$a^x=(\rho e^{i\varphi})^x=\rho^x e^{i\varphi x}=$$
$$=\rho^x[\cos(\varphi x)+i\sin(\varphi x)] \quad (14.3.2)$$

разложится в произведение двух сомножителей: первый из них (он определяет модуль числа a^x) будет монотонно возрастать или монотонно убывать с ростом x, а второй, равный $e^{i\varphi x}$ (он отвечает за аргумент числа a^x), будет меняться периодически. Если $\rho<1$, то вещественная часть и коэффициент при мнимой части числа a^x, т. е. выражения $\rho^x\cos(\varphi x)$ и $\rho^x\sin(\varphi x)$, будут моделировать *затухающие колебания* (см. § 10.4, в частности рис. 10.4.2). Ниже (см. § 17.1) мы увидим, что эта связь функции a^x с затухающими колебаниями отнюдь не случайна.

Формула Эйлера (2.6) позволяет строить многие другие важные функции вещественной, мнимой или комплексной переменной. Начнем с того, что перепишем эту формулу по-другому, выразив не показательную функцию через тригонометрические, а, напротив, тригонометрические функции через показательную. Из того, что

$$e^{i\varphi}=\cos\varphi+i\sin\varphi, \quad (14.3.3)$$

следует также, что

$$e^{-i\varphi}=\cos(-\varphi)+i\sin(-\varphi)=$$
$$=\cos\varphi-i\sin\varphi. \quad (14.3.3а)$$

Сложив почленно (3) и (3а), а затем вычитая (3а) из (3), мы без труда получим

$$\cos\varphi=\frac{e^{i\varphi}+e^{-i\varphi}}{2},\quad \sin\varphi=\frac{e^{i\varphi}-e^{-i\varphi}}{2i}. \quad (14.3.4)$$

Следствия (4) формулы (3) часто также называют формулами Эйлера.

Заметим, что при вещественном φ стоящие в правой части (4) выражения, разумеется, вещественны (это ведь суть косинус и синус некоторого угла!): в самом деле, подставив в эти выражения вместо $e^{i\varphi}$ и $e^{-i\varphi}$ разложение (2.8) и аналогичное, мы придем к знакомым уже нам формулам (2.9) и (2.9а)! Можно и по-другому с одного взгляда убедиться в вещественности полученных выражений. Заменим в выражении $z=P+iQ$ для комплексного числа z (P и Q вещественны) i на $-i$; мы придем к комплексно-сопряженному числу $z^*=P-iQ$. Если же при этом z не изменилось, т. е. $z^*=z$, то $Q=0$ и z вещественно. Но легко видеть, что (при вещественном φ) правые части (4) не меняются при замене i на $-i$; поэтому они вещественны.

Далее с помощью выражений (4) для $\cos\varphi$ и $\sin\varphi$ через экспоненты можно определить и тригонометрические функции мнимого аргумента. Подставим в (4) $\varphi=i\psi$, полагая ψ вещественным, так что φ — *чисто мнимое*. Получим

$$\cos(i\psi)=\frac{e^{-\psi}+e^{\psi}}{2},\quad \sin(i\psi)=$$
$$=\frac{e^{-\psi}-e^{\psi}}{2i}=i\,\frac{e^{\psi}-e^{-\psi}}{2}. \quad (14.3.5)$$

Таким образом, мы вернулись к экспоненциальным функциям вещественного аргумента ψ, точнее — к некоторым комбинациям экспонент, о которых подробнее скажем в следующем параграфе этой главы.

[1] Мы ограничились одними лишь натуральными (целыми положительными) значениями показателя степени; позволив n принимать также целые отрицательные значения и нуль, мы неограниченно продолжим периодическую последовательность (1) чисел $(-1)^n$ влево.

Рассмотрение функций $\cos(i\psi)$ и $\sin(i\psi)$ решает и еще одну задачу. Возьмем, к примеру, $\psi=1$. Простой расчет дает

$$\cos i = \frac{e^1+e^{-1}}{2} \approx \frac{2,72+0,36}{2} = 1,54 \tag{14.3.6}$$

$\left(\text{и } \sin i \approx i\,\frac{2,72-0,36}{2} = 1,18i\right)$. Поставим вопрос: каков тот угол, косинус которого равен 1,5? Раньше мы ответили бы, что такого угла нет, ибо косинус любого угла не превосходит единицу. При этом молчаливо подразумевается, что угол, от которого зависит косинус, — это «настоящий» угол, тот, который можно нарисовать на чертеже и измерить транспортиром. Но теперь, после введения мнимых углов, можно дать определенный численный ответ: уравнение

$$\cos\varphi = 1,5 \tag{14.3.7}$$

имеет (приближенное) решение $\varphi\approx i$. Очевидно, что значение $\varphi\approx -i$ также является решением уравнения (7); наряду с этим здесь имеется еще бесконечное число решений, отличающихся одно от другого на числа, кратные 2π. Таким образом, если ранее мы не имели ни одного решения уравнения (7), то теперь можем считать, что их существует б е с к о н е ч н о   м н о г о:

$$\varphi = \pm i + 2k\pi, \text{ где } k=0,\ \pm1,\ \pm2,\ \ldots \tag{14.3.8}$$

Здесь положение вполне аналогично тому, с каким мы столкнулись в случае алгебраических уравнений. При использовании одних только вещественных чисел квадратное уравнение $x^2-n=0$, где $n>0$, имело два решения (корня) $x=\pm\sqrt{n}$, а уравнение $x^2+n=0$ — ни одного решения; при переходе же к комплексным числам второе уравнение также получало два решения: $x=\pm i\sqrt{n}$. Аналогично уравнение $\cos\varphi=n$ при $n\leqslant 1$ имело бесконечно много решений (так, значению $n=1/2$ отвечают углы $\varphi=\pm\frac{\pi}{3}+2k\pi$, где k — любое целое число), а при $n>1$ оно не имело ни одного решения; при переходе же к комплексным значениям φ и это уравнение приобретает бесконечное множество решений.

Задачу о решении уравнений можно сформулировать как вопрос об определении о б р а т н о й   ф у н к ц и и. Рассмотрим («прямую») функцию $t=x^2$; задача нахождения обратной функции сводится к решению уравнения $x^2=t$, где t задано, а x надо найти. При этом в области вещественных переменных исходная функция существует при всех значениях аргумента x, а обратная — лишь при значениях аргумента $t\geqslant 0$; в комплексной же плоскости обратная функция также определена при в с е х t. Аналогично и функция $\varphi=\arccos s$, обратная функции $s=\cos\varphi$, в области вещественных переменных существует лишь при $-1\leqslant s\leqslant 1$ (ср. с § 4.11); в области же комплексных переменных она определена при в с е х s.

Оказывается, что так же обстоит дело и с   л о г а р и ф м и ч е с к о й   функцией $z=\ln w$, которая определяется как обратная показательной функции $w=e^z$. Мы знаем, что в области вещественных чисел логарифмы существуют не всегда: так, например, не имеют логарифмов отрицательные числа. В комплексной же области, как мы убедимся чуть ниже, логарифмы определены всегда. Наконец, заметим, что все три обратные функции: $x=\sqrt{t}$, $\varphi=\arccos s$ и $z=\ln w$ — являются неоднозначными, хотя характер этой неоднозначности здесь неодинаков; на причинах, обусловливающих это различие, мы еще остановимся ниже.

Перепишем формулу (2.10) так:

$$e^{x+iy} = \rho(\cos y + i\sin y) = \rho e^{iy},$$

$$\text{где } \rho=e^x, \text{ т. е. } x=\ln\rho. \tag{14.3.9}$$

Поскольку равенство $w=e^z$ равносильно записи $z=\ln w$, то из (9) вытекает: если $w=\rho(\cos\varphi+i\sin\varphi)$, то

$$z=\ln w = \ln\rho + i\varphi = \ln|w| + i\operatorname{Arg} w. \tag{14.3.10}$$

Разумеется, соотношение (10) надо считать о п р е д е л е н и е м (ранее нам неизвестного) числа $\ln w$: ведь «вывод» равенства (10) из (9) мы мотивировали равносильностью записей $e^z=w$ и $z=\ln w$, справедливой для вещественных z и w, но не для комплексных (или хотя бы отрицательных вещественных) w, для которых величина $\ln w$ ранее не имела смысла. Целесообразность определения (10) связана с тем, что при $\varphi=\operatorname{Arg} z=0$, т. е. в том случае, когда z — в е щ е с т в е н н о е   п о л о-

жительное число, формула (10) приводит к обычному понятию логарифма (при введении логарифмов комплексных чисел надо доучиваться — не переучиваться); кроме того, в силу (9) здесь всегда (для любого w) $e^{\ln w}=w$.

Выражая числа ρ (т. е. $|w|$) и φ (т. е. $\operatorname{Arg} w$) через вещественную часть и коэффициент при мнимой части комплексного числа $w=u+iv$: $\rho=|w|=\sqrt{u^2+v^2}$, $\varphi=\operatorname{Arg} w=\operatorname{arc\,tg}\left(\frac{v}{u}\right)$, перепишем (10) еще так:

$$\ln(u+iv)=\ln\sqrt{u^2+v^2}+i\operatorname{arc\,tg}\left(\frac{v}{u}\right). \tag{14.3.11}$$

Окончательно мы пришли к новому понятию логарифма числа — более широкому, чем известное нам из средней школы. Формула (10) (или (11)) позволяет найти, например, логарифм отрицательного числа. В самом деле, можно написать, скажем,

$$-5=5\cdot(-1)=5(\cos\pi+i\sin\pi)=5e^{i\pi},$$

откуда следует, что

$$\ln(-5)=\ln 5+i\pi\approx 1{,}6+3{,}14i. \tag{14.3.12}$$

А так как (ср. § 4.9) $\lg_{10}x=\frac{\ln x}{\ln 10}\approx\frac{\ln x}{2{,}3}$, то

$$\lg_{10}(-5)\approx(1/2{,}3)\ln(-5)\approx 0{,}7+1{,}37i.$$

Если модуль ρ комплексного числа z равен единице, т. е. если $z=\cos\varphi+i\sin\varphi$, то $\ln z=\varphi i=i\operatorname{Arg} z$. Именно эта связь между логарифмами комплексных чисел и их аргументами и обусловливает близость свойств аргументов чисел к свойствам логарифмов, заключающуюся в том, что для любых двух (комплексных) чисел u и v

$$\log(uv)=\log u+\log v,$$

$$\log\left(\frac{u}{v}\right)=\log u-\log v,$$

$$\operatorname{Arg}(uv)=\operatorname{Arg} u+\operatorname{Arg} v,$$

$$\operatorname{Arg}\left(\frac{u}{v}\right)=\operatorname{Arg} u-\operatorname{Arg} v,$$

т. е. при перемножении комплексных чисел их аргументы (как и логарифмы) складываются, а при делении вычитаются. (Под символом $\log u$ и здесь можно, разумеется, понимать логарифм по любому основанию — ведь все логарифмы пропорциональны друг другу.)

Заметим, что во всех формулах, в которые входит угол φ — аргумент комплексного числа, его можно заменить на угол $\varphi+2k\pi$, где k — любое целое число. Поэтому логарифм каждого числа имеет не одно, а бесконечно много значений. Так, в формуле (12) полный ответ имеет вид

$$\ln(-5)=\ln 5+i(\pi+2k\pi),$$
$$k=0,\ \pm1,\ \pm2,\ \ldots$$

Но то же самое относится и к логарифму любого положительного числа. Например,

$$\ln 5=1{,}6+i\cdot 2k\pi,\quad k=0,\ \pm1,\ \pm2,\ \ldots$$

или $\ln 1=i\cdot 2k\pi,\quad k=0,\ \pm1,\ \pm2,\ldots,$

так как

$$1=\cos 0+i\sin 0=\cos 2\pi+i\sin 2\pi=\ldots$$
$$\ldots=e^0=e^{2\pi i}=e^{4\pi i}=\ \ldots$$

Ниже мы еще вернемся к вопросу о том, почему логарифм не однозначен, каковы глубокие причины того, что каждое число имеет не один, а бесконечно много логарифмов.

Укажем, наконец, что понятие логарифма позволяет говорить о возведении в любую (комплексную) степень какого угодно (комплексного) числа. Так, например, попробуем определить число i^i (что получится, если число i умножить само на себя i раз?). Очевидно, имеем

$$i=\cos\frac{\pi}{2}+i\sin\frac{\pi}{2}=e^{\frac{\pi}{2}i}$$

и, значит,

$$i^i=\left(e^{\frac{\pi}{2}i}\right)^i=e^{\left(\frac{\pi}{2}i\right)i}=e^{-\frac{\pi}{2}}\approx$$
$$\approx 2{,}7^{-1{,}57}=\frac{1}{2{,}7^{1{,}57}}\approx 0{,}21$$

(это число оказалось даже вещественным!).

Упражнения

14.3.1. Пусть $u=\rho(\cos\varphi+i\sin\varphi)$, а $z=x+iy$. Выпишите все значения величины u^z.

14.3.2 (задача-шутка). Что больше: e^e или i^i? e^i или i^e?

§ 4*. Тригонометрические функции мнимого аргумента. Гиперболические функции

Вернемся снова к формулам (3.5) для тригонометрических функций мнимого аргумента:

$$\cos(i\psi) = \frac{e^{\psi} + e^{-\psi}}{2},$$

$$\sin(i\psi) = i\,\frac{e^{\psi} - e^{-\psi}}{2}. \qquad (14.4.1)$$

Входящие в эти формулы комбинации экспонент e^{ψ} и $e^{-\psi}$ имеют специальные названия и обозначения. Первая из них называется *гиперболическим косинусом*, а вторая — *гиперболическим синусом* аргумента ψ; обозначаются они так:

$$\operatorname{ch}\psi = \frac{e^{\psi} + e^{-\psi}}{2}, \quad \operatorname{sh}\psi = \frac{e^{\psi} - e^{-\psi}}{2}. \qquad (14.4.2)$$

(Буква h в написании этих функций напоминает нам о гиперболе (hyperbola по-латыни).) Таким образом,

$$\cos(i\psi) = \operatorname{ch}\psi, \quad \sin(i\psi) = i\operatorname{sh}\psi. \qquad (14.4.3)$$

Многие свойства гиперболических функций как бы копируют свойства (обыкновенных, или круговых) тригонометрических функций, иногда, впрочем, воспроизводя их в несколько искаженном виде. Так, из формул (2) следует, что $\operatorname{ch}\psi$ есть функция четная: $\operatorname{ch}(-\psi) = \frac{e^{-\psi} + e^{\psi}}{2} = \operatorname{ch}\psi$, в то время как $\operatorname{sh}\psi$ — функция нечетная: $\operatorname{sh}(-\psi) = \frac{e^{-\psi} - e^{\psi}}{2} = -\operatorname{sh}\psi$. С другой стороны, очевидно, $\operatorname{ch}(0)=1$, а $\operatorname{sh}(0)=0$ (ср. изображенные на рис. 1 графики гиперболических синуса и косинуса). На рис. 1 представлен также график *гиперболического тангенса*:

$$\operatorname{th}\psi = \frac{\operatorname{sh}\psi}{\operatorname{ch}\psi} = \frac{e^{\psi} - e^{-\psi}}{e^{\psi} + e^{-\psi}};$$

ясно, что функция $\operatorname{th}\psi$ нечетная ($\operatorname{th}(-\psi) = -\operatorname{th}\psi$) и $\operatorname{th}(0)=0$ [1]. Укажем еще, что известной формуле $\cos^2\varphi + \sin^2\varphi = 1$ отвечает следующая формула «гиперболической тригонометрии»:

$$\operatorname{ch}^2\psi - \operatorname{sh}^2\psi = 1. \qquad (14.4.4)$$

В самом деле,

$$\operatorname{ch}^2\psi - \operatorname{sh}^2\psi = \left(\frac{e^{\psi} + e^{-\psi}}{2}\right)^2 - \left(\frac{e^{\psi} - e^{-\psi}}{2}\right)^2 = \frac{e^{2\psi} + 2 + e^{-2\psi}}{4} - \frac{e^{2\psi} - 2 + e^{-2\psi}}{4} = 1.$$

Из разложений в ряд

$$e^{\psi} = 1 + \frac{\psi}{1} + \frac{\psi^2}{2!} + \frac{\psi^3}{3!} + \frac{\psi^4}{4!} + \ldots,$$

$$e^{-\psi} = 1 - \frac{\psi}{1} + \frac{\psi^2}{2!} - \frac{\psi^3}{3!} + \frac{\psi^4}{4!} + \ldots$$

и формул (2) следует, что

$$\operatorname{ch}\psi = 1 + \frac{\psi^2}{2!} + \frac{\psi^4}{4!} + \frac{\psi^6}{6!} + \ldots,$$

$$\operatorname{sh}\psi = \psi + \frac{\psi^3}{3!} + \frac{\psi^5}{5!} + \frac{\psi^7}{7!} + \ldots \qquad (14.4.5)$$

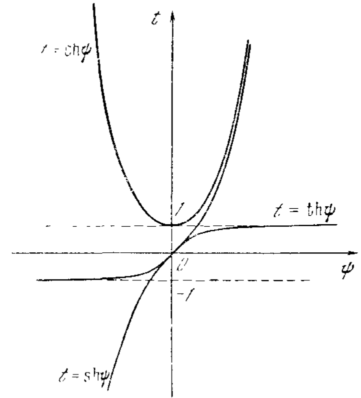


Рис. 14.4.1

(ср. с разложениями (2.9) и (2.9а) в ряд тригонометрических функций), откуда, в частности, вытекает, что при малых ψ

$$\operatorname{sh}\psi \approx \psi, \quad \operatorname{ch}\psi \approx 1 + \frac{\psi^2}{2}, \quad \operatorname{th}\psi \approx \psi \qquad (14.4.6)$$

(во всех формулах (6) мы пренебрегли членами третьего и более высоких порядков малости относительно ψ). Далее, очевидно,

$$(\operatorname{sh}\psi)' = \left(\frac{e^{\psi} - e^{-\psi}}{2}\right)' = \frac{e^{\psi} + e^{-\psi}}{2} = \operatorname{ch}\psi, \qquad (14.4.7)$$

$$(\operatorname{ch}\psi)' = \left(\frac{e^{\psi} + e^{-\psi}}{2}\right)' = \frac{e^{\psi} - e^{-\psi}}{2} = \operatorname{sh}\psi,$$

откуда в свою очередь с учетом (4) вытекает, что

$$(\operatorname{th}\psi)' = \left(\frac{\operatorname{sh}\psi}{\operatorname{ch}\psi}\right)' = \frac{(\operatorname{sh}\psi)'\operatorname{ch}\psi - \operatorname{sh}\psi(\operatorname{ch}\psi)'}{\operatorname{ch}^2\psi} = \frac{\operatorname{ch}^2\psi - \operatorname{sh}^2\psi}{\operatorname{ch}^2\psi} = \frac{1}{\operatorname{ch}^2\psi}$$

(ср. с формулами производных для тригонометрических функций § 4.10).

[1] Мы знаем, что (обычные) косинус и синус (вещественного переменного) ограничены: $|\cos\varphi| \leqslant 1$, $|\sin\varphi| \leqslant 1$, в то время как тангенс может принимать сколь угодно большие значения. Гиперболические же косинус и синус не ограничены, ибо, как легко видеть, $\operatorname{ch}\psi \to \infty$ при $\psi \to \infty$, $\operatorname{sh}\psi \to \infty$ при $\psi \to \infty$; напротив, $\operatorname{th}\psi$ ограничен: $|\operatorname{th}\psi| < 1$ (и, как нетрудно понять, $\lim_{\psi\to\infty}\operatorname{th}\psi = 1$).

Еще более удивляет сходство между известными («школьными») формулами сложения для тригонометрических функций

$$\sin(\alpha+\beta)=\sin\alpha\cos\beta+\cos\alpha\sin\beta,$$

$$\cos(\alpha+\beta)=\cos\alpha\cos\beta-\sin\alpha\sin\beta,$$

$$\operatorname{tg}(\alpha+\beta)=\frac{\operatorname{tg}\alpha+\operatorname{tg}\beta}{1-\operatorname{tg}\alpha\operatorname{tg}\beta}$$

и аналогичными формулами для гиперболических функций

$$\begin{aligned}\operatorname{sh}(u+v)&=\operatorname{sh}u\operatorname{ch}v+\operatorname{ch}u\operatorname{sh}v,\\ \operatorname{ch}(u+v)&=\operatorname{ch}u\operatorname{ch}v+\operatorname{sh}u\operatorname{sh}v,\end{aligned}\qquad(14.4.8)$$

$$\operatorname{th}(u+v)=\frac{\operatorname{sh}(u+v)}{\operatorname{ch}(u+v)}=$$

$$=\frac{\operatorname{sh}u\operatorname{ch}v+\operatorname{ch}u\operatorname{sh}v}{\operatorname{ch}u\operatorname{ch}v+\operatorname{sh}u\operatorname{sh}v}=$$

$$=\frac{\operatorname{sh}u/\operatorname{ch}u+\operatorname{sh}v/\operatorname{ch}v}{1+(\operatorname{sh}u\operatorname{sh}v)/(\operatorname{ch}u\operatorname{ch}v)}=\frac{\operatorname{th}u+\operatorname{th}v}{1+\operatorname{th}u\operatorname{th}v}$$

(последнюю из формул (8) мы получили, разделив числитель и знаменатель

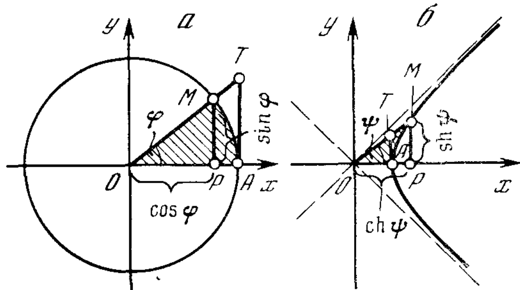


Рис. 14.4.2

дроби $\frac{\operatorname{sh}(u+v)}{\operatorname{ch}(u+v)}$ на $\operatorname{ch}u\operatorname{ch}v$; по поводу первых двух формул см. упр. 1). Из (8) в свою очередь вытекает:

$$\begin{aligned}\operatorname{sh}2u&=2\operatorname{sh}u\operatorname{ch}u,\\ \operatorname{ch}2u&=\operatorname{ch}^2u+\operatorname{sh}^2u,\\ \operatorname{th}2u&=\frac{2\operatorname{th}u}{1+\operatorname{th}^2u}\end{aligned}\qquad(14.4.9)$$

(к формулам (9) мы приходим, положив в (8) $u=v$; не правда ли, все формулы (4)—(9) очень похожи на знакомые Вам?).

Названия *гиперболических* функций ch ψ и sh ψ связаны со следующим обстоятельством. Обыкновенные (круговые) косинус и синус можно определить как абсциссу и ординату точки M единичной о к р у ж н о с т и $x^2+y^2=1$ (рис. 2, *а*); при этом аргументом здесь является угол $\widehat{AOM}=\varphi$, который можно понимать как удвоенную площадь сектора S_{AOM} круга (ведь наш круг имеет радиус 1). Аналогично гиперболические косинус и синус можно также описать как абсцису и ординату переменной точки M единичной г и п е р б о л ы $x^2-y^2=1$ (рис. 2, *б*), где в качестве аргумента выступает *гиперболический* угол ψ — удвоенная площадь сектора AOM гиперболы (см. упр. 5). Рис. 2 позволяет также геометрически представить круговой и гиперболический тангенсы:

$$\operatorname{tg}\varphi=\frac{\sin\varphi}{\cos\varphi}=\frac{MP}{OP}=\frac{AT}{OA}=AT\quad(\text{см. рис. 2, }а),$$

$$\operatorname{th}\psi=\frac{\operatorname{sh}\psi}{\operatorname{ch}\psi}=\frac{MP}{OP}=\frac{AT}{OA}=AT\quad(\text{см. рис. 2, }б)$$

(напомним, что в обоих случаях $OA=1$); отрезок AT называется *линией тангенса*, подобно тому как отрезки MP и OP — это *линии синуса* и *косинуса*. Из этого описания гиперболических функций сразу вытекают многие их свойства: и формула (4) (соотношения $\cos^2\varphi+\sin^2\varphi=1$ и $\operatorname{ch}^2\psi-\operatorname{sh}^2\psi=1$ — это не что иное, как уравнения $x^2+y^2=1$ и $x^2-y^2=1$ окружности и равносторонней гиперболы), и четность функции ch ψ, и нечетность функций sh ψ и th ψ (из рис. 2, *б* сразу видно, что $\operatorname{ch}(-\varphi)=\operatorname{ch}\psi$, а $\operatorname{sh}(-\psi)=-\operatorname{sh}\psi$ и $\operatorname{th}(-\psi)=-\operatorname{th}\psi$), и равенства $\operatorname{ch}(0)=1$, $\operatorname{sh}(0)=\operatorname{th}(0)=0$ и др.[2]

Теперь мы можем по-новому взглянуть на связь (3) между круговыми и гиперболическими функциями. Первоначально синус и косинус угла φ определялись как отрезки в круге единичного радиуса (см. рис. 2,*а*) или как отношения сторон прямоугольного треугольника с острым углом φ. Но эти определения ничего не говорят, скажем, о том, что такое $\sin i$ или $\cos i$. Что это за странный угол величины i? Как можно произвести поворот на такой угол? Построить прямоугольный треугольник с таким углом? Эти последние вопросы остаются, понятно, без ответа — но с помощью формул Эйлера найти величину выражений $\sin i$ и $\cos i$ вполне возможно. Заметьте, что ряд (2.9) для косинуса $\cos\varphi=1-\frac{\varphi^2}{2!}+\frac{\varphi^4}{4!}-\dots$

[2] Заметим, что также и формулы сложения (8) или определения (2) гиперболических функций могут быть выведены из указанного на рис. 2, *б* геометрического введения гиперболических функций (см., например: *Шерватов В. Г.* Гиперболические функции. М.: Физматгиз, 1958).

состоит из одних лишь четных степеней угла φ; поэтому при подстановке $\varphi=i$ (или, общее, $\varphi=\psi i$, где ψ — вещественное число) все члены ряда останутся вещественными: косинус чисто мнимого угла является вещественным числом. Ряд (2.9а) для синуса $\left(\sin\varphi=\varphi-\frac{\varphi^3}{3!}+\ldots\right)$ при подстановке $\varphi=i\psi$ дает чисто мнимое выражение, откуда и возникает коэффициент i в формуле (3), связывающей $\sin(i\psi)$ с вещественной функцией $\operatorname{sh}\psi$.

Аналогия между тригонометрическими и гиперболическими функциями находит дальнейшее продолжение в важной теме о дифференциальных уравнениях. Мы знаем, что дифференциальное уравнение

$$\frac{d^2x}{dt^2}=-kx \qquad (14.4.10)$$

при $k>0$ допускает общее решение

$$x=A\cos\omega t+B\sin\omega t,\quad \omega=\sqrt{k} \qquad (14.4.11)$$

(см. § 10.2). С другой стороны, уравнение

$$\frac{d^2x}{dt^2}=kx, \qquad (14.4.10а)$$

где тоже $k>0$, как мы уже видели выше (ср. § 13.10), имеет частные решения $x=e^{\sqrt{k}t}$ и $e^{-\sqrt{k}t}$, так что общее его решение можно записать в виде $x=Ce^{\sqrt{k}t}+De^{-\sqrt{k}t}$,

где постоянные C и D произвольны. Переписав это решение в виде

$$x=(C+D)\frac{e^{\sqrt{k}t}+e^{-\sqrt{k}t}}{2}+$$
$$+(C-D)\frac{e^{\sqrt{k}t}-e^{-\sqrt{k}t}}{2}$$

и обозначив $C+D=A$, $C-D=B$, мы в силу определений (2) придем к записи общего решения уравнения (10а), во всем аналогичной виду (11) общего решения уравнения (10):

$$x=A\operatorname{ch}\omega t+B\operatorname{sh}\omega t,\quad \text{где}\ \omega=\sqrt{k}. \qquad (14.4.11а)$$

Связь между тригонометрическими и показательной (или гиперболическими) функциями проливает новый свет на возникновение в сходных условиях периодических движений (колебаний) и апериодических движений. Рассмотрим шарик, находящийся в состоянии неустойчивого равновесия на круглой вершине холма (рис. 3,*а*). На самой вершине холма сила тяжести полностью уравновешивается сопротивлением опоры (холма), т. е. составляющая силы тяжести, направленная по касательной к поверхности холма и стремящаяся сдвинуть шарик с места, равна точно нулю. Однако стоит шарику немного сдвинуться, скажем вправо, — и на него начинает действовать сила F в направлении, увеличивающем этот сдвиг, причем эта сила растет с увеличением

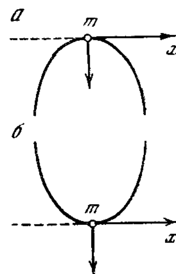


Рис. 14.4.3

сдвига x (можно, например, считать силу пропорциональной сдвигу: $F=kx$, где $k>0$). По второму закону Ньютона (см. § 9.4) уравнение движения шарика будет таково:

$$m\frac{d^2x}{dt^2}=F=kx. \qquad (14.4.12)$$

Решение этого уравнения (ср. с уравнением (10а)) имеет вид

$$x=ae^{\sqrt{\frac{k}{m}}t}+be^{-\sqrt{\frac{k}{m}}t}, \qquad (14.4.13)$$

или, что то же самое,

$$x=A\operatorname{ch}\omega t+B\operatorname{sh}\omega t,\quad \text{где}\ \omega=\sqrt{\frac{k}{m}}. \qquad (14.4.14)$$

Наличие экспоненциально растущего решения[3] $y_1=e^{\sqrt{\frac{k}{m}}t}$ означает *неустойчивость* состояния равновесия шарика на вершине холма. Можно ожидать, что в общем случае начальные условия та-

[3] Или (также экспоненциально растущих!) решений $Y_1=\operatorname{ch}\omega t$, $Y_2=\operatorname{sh}\omega t$. Легко понять, что с ростом ψ функции $u=\operatorname{ch}\psi$ и $v=\operatorname{sh}\psi$ растут как экспонента: $\lim\limits_{\psi\to\infty}\frac{\operatorname{ch}\psi}{e^\psi}=\lim\limits_{\psi\to\infty}\frac{\operatorname{sh}\psi}{e^\psi}=\frac{1}{2}$.

ковы, что в (13) $a \neq 0$ и $b \neq 0$ (было бы весьма маловероятно, чтобы начальные условия гарантировали строгое соблюдение равенства $a=0$). Следовательно, через некоторое время первое слагаемое правой части (13) обязательно станет достаточно большим — шарик упадет.

Теперь обратимся к другой, в определенном смысле противоположной ситуации — пусть шарик находится в яме с покатыми краями (рис. 3, *б*). Когда шарик сдвигается на расстояние x из положения равновесия, возникает сила F, направленная в сторону нижней точки ямы, — «возвращающая» сила. (Такое направление силы согласуется с тем, что равновесие в случае ямы *устойчиво* в отличие от равновесия шарика на горе.) Предположим, что сила F по величине пропорциональна отклонению x, так что $F=-kx$, где k — положительная величина, а знак минус в формуле как раз и говорит о том, что F стремится вернуть шарик в исходное положение. Тогда уравнение движения шарика будет таково:

$$m\frac{d^2x}{dt^2}=-kx. \qquad (14.4.12а)$$

Формально это уравнение похоже на предыдущее. Поэтому можно взять решение уравнения (12), заменив в нем k на $-k$, — ведь именно эта подстановка переводит уравнение (12) в уравнение (12а)! При этом мы получим решение новой задачи:

$$x=ae^{\sqrt{-\frac{k}{m}}\,t}+be^{-\sqrt{-\frac{k}{m}}\,t}=ae^{i\omega t}+ \\ +be^{-i\omega t}, \quad \omega=\sqrt{\frac{k}{m}}, \qquad (14.4.13а)$$

или, при другом выборе произвольных постоянных,

$$y=A\cos\omega t+B\sin\omega t, \quad \omega=\sqrt{\frac{k}{m}} \qquad (14.4.14а)$$

(ср. (11); выражения (13а) и (14а) совпадают, если положить $A=\frac{a+b}{2}$, $B=\frac{(a-b)\,i}{2}$). При этом (вещественное!) решение (14а) уравнения (12а) является уже ограниченным при всех t, что тесно связано с устойчивостью состояния равновесия «в яме».

Мнимые «числа» (как корни квадратные из отрицательных чисел) впервые (причем довольно формально) ввел видный итальянский математик эпохи Возрождения Джироламо Кардано (1501—1576). Появление комплексных чисел именно в этот период бурного развития алгебры (в то время как в античный период из всех разделов математики наибольшее внимание привлекала геометрия; интерес к алгебре в эпоху Возрождения подготовил создание в XVII в. математического анализа) и именно у Дж. Кардано никак не было случайностью. В истории математики имя Кардано осталось в первую очередь в связи с впервые им опубликованной [4] формулой для решения (произвольного [5]) кубического уравнения: если

$$x^3+px+q=0, \qquad (14.4.15)$$

то

$$x=\sqrt[3]{-\frac{q}{2}+\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}+ \\ +\sqrt[3]{-\frac{q}{2}-\sqrt{\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3}}. \qquad (14.4.16)$$

Но кажущаяся достаточно простой формула Кардано (16) является коварной: так, если уравнение (15) имеет вид $x^3-3x=0$ (т. е. если $p=-3$, $q=0$; ясно, что наше уравнение имеет простые корни $x_1=0$, $x_2=\sqrt{3}\approx 1{,}73$, $x_3=-\sqrt{3}\approx -1{,}73$), то (16) неожиданно дает

$$x=\sqrt[3]{\sqrt{-1}}+\sqrt[3]{-\sqrt{-1}}=\sqrt[3]{i}+\sqrt[3]{-i},$$

т. е. здесь вдруг появляются корни квадратные из числа -1.

Процедура извлечения (квадратных и кубических) корней из комплексных чисел была разработана последователем Кардано Рафаэлем Бомбелли (ок. 1530—1572), сумевшим объяснить, каким образом формула (16) во всех случаях гарантирует правильное определение (трех, хотя и не обязательно различных) корней уравнения (15) (ср. упр. 4).

Крупнейший алгебраист XVI в. француз Франсуа Виет (1540—1603) показал, что в том случае, когда $\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3<0$, т. е. когда корни третьей степени в формуле Кардано (16) приходится извлекать из комплексных чисел, стоящее в правой части (16) выражение удается свести к несложной комбинации тригонометрических функций. Этим самым были проложены первые пути к использованию

[4] Но не впервые им найденной: формулу (16) Кардано заимствовал у Никколо Тарталья (ср., например: *Шереметьевский В. П.* Очерки по истории математики. М.: Учпедгиз, 1940, с. 89—94).

[5] Относительно сведения любого кубического уравнения $ax^3+bx^2+cx+d=0$ к виду (15) см. с. 35.

тригонометрических функций в учении о возведении комплексных чисел в степени (в том числе — в отрицательные и дробные степени), т. е. к общей формуле (1.6). Однако в полном виде (1.6) впервые выписал лишь математик XVII в. Абрахам де Муавр (1667—1754); современный же вид этой формуле придал Л. Эйлер. (Француз по национальности де Муавр вынужден был покинуть Францию из-за гонений на протестантов (гугенотов); он переселился в Англию и сыграл большую роль в развитии математики в этой стране.)

Весьма широко использовал комплексные числа знаменитый Леонард Эйлер. Ему принадлежит само обозначение i для «числа» $\sqrt{-1}$; от него же идут почти все результаты настоящей главы. Эйлер был уверен в справедливости основной теоремы алгебры, утверждающей существование у каждого алгебраического уравнения n-й степени (с вещественными или комплексными коэффициентами) n (вообще говоря, комплексных) корней, и многократно пытался доказать ее, что, однако, требовало полного понимания алгебраического и геометрического (см. рис. 1.1) смысла комплексных чисел, которого в тот период у математиков еще не было. Ближе подошел к доказательству основной теоремы алгебры второй крупнейший математик XVIII столетия Жан Лерон Д'Аламбер — однако и он не довел ее доказательство до конца.

Первым доказательством (или, точнее, первыми доказательствами — их было несколько) теоремы о существовании у любого уравнения n-й степени ровно n (вещественных или комплексных) корней мы обязаны Карлу Фридриху Гауссу. От Гаусса же идет современное представление плоскости комплексной переменной $z=x+iy$, снявшее с комплексных чисел последний налет загадочности и даже мистики [6]; ему принадлежат также многие другие глубокие результаты в области «комплексной алгебры и анализа». В частности, Гаусс является создателем нетривиальной «арифметики целых комплексных чисел», т. е. чисел вида $a+ib$, где a и b — целые; в этой необыкновенной арифметике число 3 по-прежнему считается *простым*, но число 5 уже является *составным*, поскольку $5=(1+2i)\times\times(1-2i)$. Можно сказать, что теория комплексных чисел стала полноправным разделом математики только начиная с Гаусса. (Геометрическое истолкование комплексных чисел (см. рис. 1.1) было до Гаусса дано датским землемером и математиком-любителем Каспером Весселем (1745—1818) и швейцарским ученым Жаном Робертом Арганом (1768—1822) — но публикации мало известных Весселя и Аргана не были своевременно замечены, и Гаусс пришел к этим идеям самостоятельно.)

[6] Но ведь и гораздо более простые отрицательные числа были впервые введены в науку лишь Дж. Кардано и Ф. Виетом, причем еще даже и во времена Виета математики относились к ним весьма недоверчиво!

Упражнения

14.4.1. Докажите формулы (8) сложения для гиперболических функций.

14.4.2. Докажите, что $\operatorname{sh} t=\dfrac{2\operatorname{th}(t/2)}{1-\operatorname{th}^2(t/2)}$,

$$\operatorname{ch} t=\frac{1+\operatorname{th}^2(t/2)}{1-\operatorname{th}^2(t/2)},\quad \operatorname{th} t=\frac{2\operatorname{th}(t/2)}{1+\operatorname{th}^2(t/2)}.$$

14.4.3. Какие линии задаются параметрическими уравнениями:

а) $x=\dfrac{1-t^2}{1+t^2},\quad y=\dfrac{2t}{1+t^2}$;

б) $x=\dfrac{1+t^2}{1-t^2},\quad y=\dfrac{2t}{1-t^2}$?

Какой геометрический смысл имеет параметр t в этих формулах?

14.4.4. Поясните, как из формулы (16) возникают значения 0 и $\pm\sqrt{3}$ корней рассмотренного в тексте кубического уравнения.

14.4.5. Докажите, что в обозначениях рис. 2, *б* действительно $\operatorname{ch}\psi=OP$ и $\operatorname{sh}\psi=MP$, где ψ — удвоенная площадь заштрихованного на этом рисунке гиперболического сектора.

Глава 15

КАКИЕ ФУНКЦИИ НУЖНЫ ФИЗИКУ?

§ 1. Аналитические функции вещественной переменной

Понятие функции и функциональной зависимости претерпело большие изменения в ходе развития математики. Различный подход к функции в разные эпохи и со стороны разных ученых, в иные эпохи приводивший к бурным дискуссиям [1], отражается еще и сегодня в том, что функцию можно определять по-разному.

Самое широкое определение состоит в том, что любое соотношение элементов двух множеств можно назвать «функцией». Если Маша носит белое платье, Даша — красное, Наташа — черное, то можно сказать, что цвет платья является функцией имени девушки (или платье является функцией определенного лица); здесь область определения функции («область отправления» сотнесения, или соответствия, как стали говорить математики в XX в.) состоит из трех имен (или трех девушек): Маша, Даша, Наташа, — а область ее значений («область прибытия») — из трех цветов: белый, красный, черный (или из трех платьев). Функцией является и экспериментально установленная зависимость одних физических величин от других, например электрического сопротивления проволоки от температуры. Наконец, функцией, несомненно, является зависимость, заданная формулой, например $y=ax^2+bx+c$ или $y=e^{-x^2/2}$, — и подобных формул можно написать очень много!

Полезно ли объединять в одну категорию и называть одним словом все эти очень разнородные понятия? В некоторых отношениях это полезно, так как подчеркивает огромную общность понятия функции. Но общность эта дается отнюдь не даром!

Функцию, заданную чисто словесно (цвет платья как функция имени девушки), нельзя выразить формулой, нельзя найти производную или интеграл такой функции — весь разработанный в течение столетий аппарат высшей математики оказывается бесполезным, ибо к таким функциям он неприменим.

Функция, полученная на опыте и заданная в виде таблицы чисел, всегда оказывается заданной лишь для определенного набора значений аргумента и определена с некоторой погрешностью [2]. Можно ли вычислить производную или интеграл такой функции? По смыслу производной (и интеграла) вычисление связано с определением значений функции в сколь угодно близких точках, при сколь угодно близких значениях независимой переменной. Непосредственно из опыта или из таблицы такие значения функции получить нельзя, особенно если учесть еще ошибки измерения. Надо предположить, что существует плавная функциональная зависимость между измеряемыми величинами — только тогда можно использовать аппарат высшей математики; данные же таблицы или опыта нам нужны для того, чтобы эту зависимость найти.

Огромное значение имеет тот факт, что фундаментальная теория всегда приводит к достаточно естественным математическим формулам. Это справедливо и тогда, когда нам приходится обращать те или иные зависимости (например, решать алгебраические уравнения) и при этом получаются решения с разрывами, переходами от одного корня уравнения к другому. Физическая теория всегда строится так, чтобы сохранить возможность применения столь бесценного аппарата исследования явлений,

---

[1] Наиболее известен здесь упоминавшийся выше (см. § 10.8) спор на эту тему между тремя выдающимися математиками XVIII в. — Л. Эйлером, Ж. Л. Д'Аламбером и Д. Бернулли (каждый из участников которого так и остался при своем первоначальном мнении): речь шла о том, можно ли считать единой функцией зависимость, график которой претерпевает разрыв или имеет излом.

[2] С неизбежными погрешностями мы встречаемся и в том случае, когда функция с самого начала задается своим графиком (например, вычерчивается самопишущим устройством измерительного прибора).

как дифференциальное и интегральное исчисления.

Итак, функции, заданные формулами, обладают огромными преимуществами. Прежде всего ясно, что задание формулы позволяет в принципе вычислить функцию с любой точностью при любом значении независимой переменной. В отличие от определения на опыте вычисление по формуле при повторении всегда дает одно и то же значение; точность вычисления можно заранее обусловить. Очень удобно также то, что вычисление значений аналитически заданной функции при разных значениях аргумента сводится к последовательности однотипных операций; особенно важно последнее обстоятельство при пользовании компьютерами. Задание формулы позволяет также вычислить с любой точностью разность значений функции при разных (но близких) значениях независимой переменной и таким образом численно найти производную. Суммируя значения функции в разных точках, мы можем вычислить ее интеграл; при этом и производная, и интеграл могут быть определены с любой степенью точности. В современных условиях широкого распространения компьютеров подобное прямое вычисление производной или интеграла является делом достаточно простым.

Однако и среди тех функций, для которых указан точный метод вычисления величины $f(x)$ по заданному значению x, имеются функции разной природы. Пусть, например, $f(x)=1$ или 0, причем $f(x)=1$ для всех x, выражаемых несократимыми дробями с четными знаменателями, и только для таких x [3]:

$$f(x)=\begin{cases} 1, \text{ если } x=\frac{m}{n}, \text{ где } m \text{ и } n \text{ — взаимно просты и } n \text{ — четно;} \\ 0 \text{ во всех остальных случаях.} \end{cases}$$

Попробуйте построить график такой функции — Вам никогда не удастся это сделать, ибо здесь надо будет соорудить частокол из вертикальных отрезков длины 1, восстановленных в точках оси x вида $x=\frac{m}{n}=\frac{m}{2n_1}$ (где m и n_1 — взаимно простые целые положительные числа и m — нечетно) — а таких точек на любом отрезке оси x будет бесконечно много. Разумеется, к этой функции развитый в настоящей книге аппарат тоже неприложим.

С другой стороны, мы имеем, скажем, функцию $y=1-x$ — «идеальную» функцию, изображаемую гладкой линией (даже прямой), в каждой точке имеющей касательную (она совпадает с самой линией), т. е. всюду имеющей производную; легко вычислить здесь и обратную функцию (в этом случае она совпадает с исходной) или, скажем, функцию $f(f(x))$ (это есть функция $F(x)=x$); ничего не стоит найти и интеграл от нашей функции:

$$\int_a^b f(x)dx=\left(x-\frac{x^2}{2}\right)\Big|_a^b=(a-b)\left(\frac{a+b}{2}-1\right)$$

и т. д.

Но возможны и «промежуточные» функции — не столь «плохие» (или, как говорят математики, не столь «патологические»), как первая из рассмотренных, но и не столь уж «хорошие», как вторая. Как же классифицировать «хорошие» и «плохие» функции; какие функции следует считать на самом деле «хорошими»? Существенно облегчает ответ на этот вопрос переход к комплексным числам и функциям комплексной переменной — однако сначала мы попробуем ответить на него, оставаясь в области вещественных функций.

Одним из основных результатов гл. 6 этой книги явилась возможность представления функций в виде ряда

$$f(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\frac{f'''(0)}{3!}x^3+\ldots \quad (15.1.1)$$

(ряд Маклорена) или

$$f(x)=f(a)+\frac{f'(a)}{1!}(x-a)+\frac{f''(a)}{2!}(x-a)^2+\frac{f'''(a)}{3!}(x-a)^3+\ldots \quad (15.1.2)$$

(ряд Тейлора, обобщающий ряд Маклорена). Эти ряды дают возможность, зная значение функции в одной точке и ее производные в этой же точке, сколь

[3] Даже и столь «экзотическую» (но «математически», т. е. точно, заданную) функцию можно выразить (достаточно, впрочем, сложной) формулой.

угодно точно реконструировать функцию в окрестности этой точки — и затем продолжить ее далее. Если имеет место, скажем, формула (1), то, дифференцируя ее почленно, мы найдем разложение в ряд и для производной функции:

$$f'(x)=f'(0)+\frac{f''(0)}{1!}x+\frac{f'''(0)}{2!}x^2+\ldots, \tag{15.1.3}$$

а если понадобится — то и для последующих производных; аналогично из (1) следует

$$\int f(x)\,dx=C+f(0)x+\frac{f'(0)}{2!}x^2+\frac{f''(0)}{3!}x^3+\frac{f'''(0)}{4!}x^4+\ldots \tag{15.1.4}$$

К подобным функциям с наибольшей полнотой применим развитый в на-

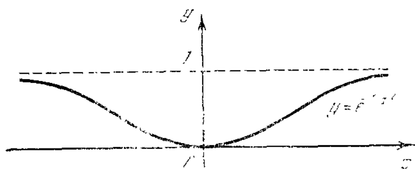


Рис. 15.1.1

стоящей книге аппарат. Представимые в виде (1) (или (2)) функции называются *аналитическими* [4].

Ясно, что каждая целая рациональная функция (многочлен)

$$f(x)=a_0+a_1x+a_2x^2+\ldots+a_nx^n \tag{15.1.5}$$

является аналитической: здесь $f(0)=a_0$, $f'(0)=a_1$, $f''(0)=2!a_2$, $f'''(0)=3!a_3, \ldots, f^{n}(0)=n!a_n$ и $f^{(m)}(0)=0$ при $m>n$ (таким образом, представление (1) такой функции совпадает с ее записью (5)). Можно сказать, что аналитические функции (подобные, например, функциям $e^x=1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\ldots$, или $\cos x=1-\frac{x^2}{2!}+\frac{x^4}{4!}-\ldots$, или $\ln x=1+x+\frac{x^2}{2}+\frac{x^3}{3}+\ldots$ (см. гл. 6)) являются наиболее естественным обобщением многочленов: общая аналитическая функция (1) (или (2)) — это, так сказать, «многочлен бесконечно большой степени».

Конечно, любая аналитическая в точке $x=a$ функция является в этой точке «беспредельно гладкой», т. е. имеет здесь производные в с е х порядков — ведь эти производные фигурируют в представлении (2), существование которого равносильно аналитичности функции. Однако в принципе из существования производных всех порядков аналитичность еще н е в ы т е к а е т: требование аналитичности функции является б о л е е с и л ь н ы м, чем требование ее неограниченной дифференцируемости. В самом деле, существование всех производных равносильно возможности выписать стоящий в правой части (2) ряд, но надо еще, чтобы существовала (при всех x, близких к a) сумма этого ряда и чтобы она была равна исходной функции $f(x)$. Проиллюстрировать сказанное можно на примере функции $y=e^{-1/x^2}$ (рис. 1). При $x=0$ выражение e^{-1/x^2} смысла не имеет; однако поскольку при малых по абсолютной величине x число $\frac{1}{x^2}$ очень велико и соответственно e^{-1/x^2} очень мало, то естественно положить, что $y(0)=0$. Нетрудно видеть, что здесь существуют все производные $y^{(n)}(x)=\frac{d^ny}{dx^n}$, где $n=1, 2, 3, \ldots$; их легко найти по формулам гл. 4 (так, например, $\frac{dy}{dx}=e^{-1/x^2}\frac{d(-1/x^2)}{dx}=\frac{2}{x^3}e^{-1/x^2}$). Однако все они при $x=0$ оказываются равными 0 (см. упр. 1) — и, значит, правая часть ряда Маклорена (1) нашей функции тождественно равна 0, в то время как сама функция, разумеется, при $x\neq 0$ в нуль не обращается. Однако, к счастью, подобные e^{-1/x^2} функции являются весьма редкими исключениями и в физических или технических задачах почти не встречаются [5].

[4] Разумеется, формулы (1) и (2) могут быть применены лишь в некоторой ограниченной области изменения x (в ограниченной окрестности значения $x=0$ или $x=a$), о чем мы подробнее говорили в § 6.3. В соответствии с этим строгий смысл имеет понятие функции, *аналитической в точке* (такой, что в окрестности рассматриваемой точки имеют место формулы (1) или (2)); функция же, *аналитическая на отрезке*, — это функция, аналитическая в к а ж д о й точке рассматриваемого отрезка. (Мы не останавливаемся здесь также на условиях возможности почленного дифференцирования или почленного интегрирования формул (1) и (2), выполняющихся, впрочем, практически во всех интересующих естествоиспытателя или инженера случаях.)

[5] Причины, вызывающие столь странные свойства рассматриваемой функции, проясняются при ее «комплексификации» (как говорят иногда математики), т. е. при расширении ее с помощью предположения о том, что аргумент x функции может принимать любые к о м п л е к с н ы е значения; по этому поводу см. § 2.

Требование аналитичности функции является весьма сильным. Так, здесь вся функция (в пределах действия формул (1) или (2)) может быть восстановлена по своим значениям на сколь угодно малом участке (по сколь угодно малой дуге графика функции): ведь знания $f(a)$ и значений $f(x)$ при x, сколь угодно близких к a, вполне достаточно для вычисления всех производных функции в одной точке a. Поэтому от общего определения функции как произвольного соответствия между значениями x и $f(x)$ аналитические функции достаточно далеки. Однако почти все функции, задаваемые простыми формулами, являются аналитическими; функции, для которых в отдельных точках нарушается аналитичность (вроде функции e^{-1/x^2}, которую мы рассмотрели выше), можно считать патологическими.

Разумеется, в физических задачах иногда могут возникать и неаналитические в отдельных точках функции (например, изменяющиеся скачком или имеющие излом); еще более удивительным «функциям» (настолько удивительным, что само слово «функция» здесь, пожалуй, уместно заключить в кавычки) будет посвящена гл. 16. Однако в подавляющем большинстве случаев те функции, с которыми встречается естествоиспытатель, являются «хорошими», т. е. аналитическими; особенно выпукло выступает это положение тогда, когда речь идет о функциях комплексной переменной, к которым мы далее и перейдем.

Упражнение

15.1.1. Пусть $f(x)=e^{-1/x^2}$ при $x\neq 0$ и $f(0)=0$. Чему равна производная $f'(0)$ этой функции в нуле? Вторая производная $f''(0)$? n-я производная $f^{(n)}(0)$?

§ 2. Производные функций комплексной переменной

Возьмем функцию w комплексной переменной z, которая сама является, естественно, комплексной величиной: $w=f(z)$, где $w=u+iv$, а $z=x+iy$; здесь числа x, y и $u=u(x, y)$, $v=v(x, y)$ — вещественные.

Если функция $f(x)$ выражается формулой, то ее производную $\frac{df}{dz}=\lim\limits_{\Delta z\to 0}\frac{f(z+\Delta z)-f(z)}{\Delta z}$ можно найти по общим правилам, совпадающим с правилами, установленными в гл. 4 для функций вещественной переменной. Вот примеры этого:

$w=z^2$, $\frac{dw}{dz}=2z$; $w=e^{kz}$, $\frac{dw}{dz}=ke^{kz}$ (здесь k может быть и комплексным); $w=\ln z$, $\frac{dw}{dz}=\frac{1}{z}$

и т. д. Количество таких примеров неисчерпаемо: можно образовывать всяческие комбинации функций — суммы (скажем, z^2+e^z), произведения (например, z^2e^z), функции от функций (например, e^{z^2}) и т. д.; при этом, скажем, $\frac{d(z^2+e^z)}{dz}=2z+e^z$, а $\frac{de^{z^2}}{dz}=2ze^{z^2}$. Вполне естественно, что все правила, относящиеся к функциям вещественной переменной, остаются справедливыми и в случае перехода к комплексной переменной: ведь эти правила основаны на общих законах алгебры, скажем на дистрибутивном законе $(a+b)\,c=ac+bc$, указывающем, как надо раскрывать скобки при вычислениях, и на некоторых родственных ему. Законы же эти не зависят от характера величин z, Δz; w, Δw. Так, например, правила дифференцирования функции $f(z)=z^2$ или произведения w_1w_2 двух функций основаны на равенствах

$$(z+\Delta z)^2=z^2+2z\Delta z+(\Delta z)^2,$$

$$\Delta(z^2)=(z+\Delta z)^2-z^2=2z\Delta z+(\Delta z)^2,$$

соответственно

$$(w_1+\Delta w_1)(w_2+\Delta w_2)=w_1w_2+w_1\Delta w_2+w_2\Delta w_1+\Delta w_1\Delta w_2,$$

$$\Delta(w_1w_2)=(w_1+\Delta w_1)(w_2+\Delta w_2)-w_1w_2=w_1\Delta w_2+w_2\Delta w_1+\Delta w_1\Delta w_2.$$

Возведение в мнимую или комплексную степень было нами определено именно таким образом, чтобы было верно (приближенное) равенство $e^{\Delta z}\approx 1+\Delta z$, где отброшены члены более высокого порядка малости, чем Δz (см. § 14.2), т. е. по формуле, лежащей в основе вычисления производной от экспоненциальной функции e^x вещественной переменной. Поэтому мы можем не сомневаться в применимости всех формул для производных к функциям

комплексной переменной: после того, как мы определили $e^{\Delta z}$ так, что $e^{\Delta z} \approx \approx 1 + \Delta z$, все дальнейшее уже строится на вполне прочном фундаменте.

Заметим, что изменение аргумента z функции $w = f(z)$ можно произвести разными способами: можно изменить только вещественную часть z: $d_1 z = d_1 x$, $d_1 y = 0$, или только мнимую часть: $d_2 z = i d_2 y$, $d_2 x = 0$, или и вещественную, и мнимую части одновременно (рис. 1). При этом по-разному изменится и функция w; но если dz мало, то отношение $\frac{dw}{dz}$, т. е. производная, во всех случаях останется одним и тем же.

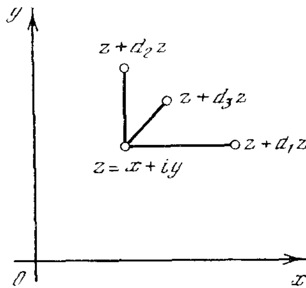


Рис. 15.2.1

Проверим это на примере функции $w = z^2$, т. е.

$$w = (x + iy)^2 = x^2 - y^2 + 2ixy = u + iv, \qquad (15.2.1)$$

где $u = x^2 - y^2$, $v = 2xy$; здесь

$$\frac{\partial u}{\partial x} = 2x, \quad \frac{\partial u}{\partial y} = -2y, \quad \frac{\partial v}{\partial x} = 2y, \quad \frac{\partial v}{\partial y} = 2x. \qquad (15.2.2)$$

Пусть $dy = 0$, т. е. $dz = dx$. Тогда изменение w равно

$$dw \approx w(z + dx) - w(z) = [u(x + dx, y) + iv(x + dx, y)] - [u(x, y) + iv(x, y)] \approx \frac{\partial u}{\partial x} dx + i \frac{\partial v}{\partial x} dx.$$

Поэтому производная в этом случае равна

$$w'(z) = \frac{dw}{dz} = \frac{dw}{dx} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = 2x + i \cdot 2y. \qquad (15.2.3)$$

Другой способ изменения z: примем $dx = 0$, $dz = i dy$; в этом случае $w + dw = w[x + (y + dy) i]$ и, значит,

$$dw = \frac{\partial u}{\partial y} dy + i \frac{\partial v}{\partial y} dy,$$

$$\frac{dw}{dz} = \frac{dw}{i dy} = -i \frac{dw}{dy} = -\frac{i \partial u}{\partial y} + \frac{\partial v}{\partial y} = -i(-2y) + 2x = 2x + i \cdot 2y. \qquad (15.2.4)$$

При преобразованиях (3) и (4) мы лишь на самом последнем этапе учли формулы (1), (2). Все же остальные рассуждения сохраняют силу для в с е х функций: если производная функции $w = w(z) = u(x, y) + iv(x, y)$ не зависит от выбора приращения dz независимой переменной, то имеет место тождество

$$\frac{dw}{dz} = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = -i \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}. \qquad (15.2.5)$$

Так как u и v вещественны, это тождество дает два соотношения — знаменитые ф о р м у л ы К о ш и — Р и м а н а:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}. \qquad (15.2.6)$$

Соотношения Коши—Римана (6) выполняются автоматически, если w задано формулой. Приведем примеры:

1) если $w = z^2 = (x + iy)^2$, то $u = x^2 - y^2$, $v = 2xy$ и $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = 2x$, $\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y} = 2y$ (ср. (2));

2) если $w = (2 + 3i) z^2 = (2 + 3i) \times [(x^2 - y^2) + i \cdot 2xy] = [2(x^2 - y^2) - 6xy] + i[3(x^2 - y^2) + 4xy]$, то $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = 4x - 6y$, $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} = -6x - 4y$;

3) если $w = e^z = e^x \cos y + ie^x \sin y$, т. е. $u = e^x \cos y$, $v = e^x \sin y$, то $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = e^x \cos y$, $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} = -e^x \sin y$.

Отметим, что условия Коши—Римана не только связывают между собой функции $u(x, y)$ и $v(x, y)$, но и накладывают определенные условия на к а ж д у ю из них. В самом деле, в силу (6)

$$\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial v}{\partial y}\right), \text{ т. е. } \frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 v}{\partial x \partial y};$$

$$\frac{\partial}{\partial y}\left(\frac{\partial v}{\partial x}\right) = \frac{\partial}{\partial y}\left(-\frac{\partial u}{\partial y}\right), \text{ т. е. } \frac{\partial^2 v}{\partial x \partial y} = -\frac{\partial^2 u}{\partial y^2}.$$

Отсюда $\frac{\partial^2 u}{\partial x^2} = -\frac{\partial^2 u}{\partial y^2}$, аналогично $\frac{\partial^2 v}{\partial x^2} = -\frac{\partial^2 v}{\partial y^2}$; другими словами,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0. \qquad (15.2.7)$$

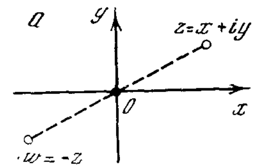
Любая функция $\varphi = \varphi(x, y)$ двух (вещественных) переменных x и y, такая, что $\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial y^2} = 0$, называется *гармонической* (ср. § 9.3). Таким образом,

если $w=f(z)=u(x, y)+iv(x, y)$ — дифференцируемая функция комплексной переменной $z=x+iy$, то вещественная и мнимая части функции w (т. е. функции $u(x, y)$ и $v(x, y)$) — гармонические функции.

Здесь уместно снова вернуться к вопросу о том, как определяется функция или как классифицируются функции, т. е. к тому вопросу, с обсуждения которого мы начали эту главу. Если речь идет о функциях комплексной переменной, то наиболее общий подход заключался бы в следующем: комплексная величина z характеризуется двумя вещественными числами x, y, или, что то же, соответствует точке на плоскости. Функция w также состоит из вещественной u и мнимой iv частей, поэтому ей тоже отвечает точка плоскости комплексной переменной (с координатами u, v). Поставим в соответствие каждому z, т. е. каждой паре x, y, или каждой точке плоскости, свои определенные u и v, которые вместе задают w (w тоже можно понимать как точку плоскости; лучше при этом говорить о д в у х плоскостях: плоскости z и плоскости w). Если задать $u(x, y)$ и $v(x, y)$ в виде гладких функций, то можно будет даже найти (частные) производные $\frac{\partial u}{\partial x}$, $\frac{\partial v}{\partial x}$, $\frac{\partial u}{\partial y}$, $\frac{\partial v}{\partial y}$; $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y}$. Однако и при этом мы, вообще говоря, не получим определенного значения производной $\frac{dw}{dz}=\frac{du+idv}{dx+idy}$ — эта величина будет зависеть от выбора dx и dy. Таким образом, общее соответствие, при котором в каждой точке $z=x+iy$ задается свое $w=u(x, y)+iv(x, y)$, не позволяет определить величину $\frac{dw}{dz}$.

Если мы пишем формулу, выражающую w как функцию z, такую, как z^2, или e^z, или $z^3e^{z^2}$, и т. д., то вещественная и мнимая части w, т. е. функции $u(x, y)$ и $v(x, y)$, оказываются не независимыми: они связаны между собой условиями Коши—Римана (6) и каждая из них подчинена условиям гармоничности (7). Итак, аналитическая формула, связывающая w и z, существенно ограничивает выбор функций $u(x, y)$ и $v(x, y)$, но зато она позволяет найти производную $\frac{dw}{dz}$; помимо этого, факт именно аналитической зависимости w от z влечет и другие важные и красивые следствия. В этом случае говорят, что $w(z)$ есть *аналитическая функция комплексной переменной* z. Все простые функции, такие, как z^2, e^z и т. п., — это аналитические функции.

Надо хорошо понимать, что аналитические функции комплексной переменной возникают при задании формулой зависимости $w=f(z)$; однако вовсе не все функции (не все отображения плоскости z в плоскость w) обязательно являются аналитическими. Так, ясно, что функция $w=-z$, переводящая каждую точку z плоскости комплексной переменной



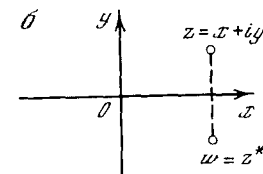


Рис. 15.2.2

в *симметричную* z *относительно начала координат* точку $w(z)$ (рис. 2, *а*), является аналитической: при этом $u=-x$, $v=-y$ и наряду с тем, что, разумеется,

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=0$$

$\left(\text{ведь в этом случае даже } \frac{\partial^2 u}{\partial x^2}=\frac{\partial^2 u}{\partial y^2}=\frac{\partial^2 v}{\partial x^2}=\frac{\partial^2 v}{\partial y^2}=0\right)$,

$$\frac{\partial u}{\partial x}=-1=\frac{\partial v}{\partial y}, \quad \text{а} \quad \frac{\partial u}{\partial y}=0=-\frac{\partial v}{\partial x}.$$

Но, скажем, функция $w=f(z)=z^*$, переводящая точку z плоскости комплексной переменной в точку $w(z)$, *симметричную* z *относительно вещественной оси* (рис. 2, *б*), аналитической уже не является. В самом деле, здесь $u=x$, $v=-y$, и хотя, конечно, по-прежнему

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=0,$$

но

$$1=\frac{\partial u}{\partial x}\neq\frac{\partial v}{\partial y}=-1, \quad \text{хотя} \quad \frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}=0$$

(а для того чтобы функция не была аналитической, достаточно, чтобы нарушалось х о т я б ы о д н о из условий Коши—Римана). Аналогично не является аналитической функ-

ция $w=|z|^2=zz^*$ (а какой алгебраической формулой, не использующей величины z^*, можно выразить зависимость $|z|^2$ от z? Да никакой!) или функция $w=f(z)=|z|=\sqrt{zz^*}$ и т. д. Так, для функции $w=|z|^2=x^2+y^2$, очевидно, $u=x^2+y^2$, $v=0$, т. е. $\frac{\partial u}{\partial x}=2x$, $\frac{\partial u}{\partial y}=2y$ и $\frac{\partial v}{\partial x}=\frac{\partial v}{\partial y}=0$, откуда и видно, что условия Коши—Римана здесь не выполняются. Вообще никакие функции комплексной переменной, записываемые с помощью величины z^*, аналитическими не являются — ни одну из них нельзя записать с помощью алгебраической (или аналитической) формулы[1]; не аналитичны также функции переменной z, в записи которых фигурирует $|z|$ или (тем более) $\operatorname{Arg} z$.

Свойства аналитических функций комплексной переменной иногда проливают новый свет на свойства обычных (вещественнозначных) функций вещественной переменной, представляющиеся зачастую загадочными без перехода к комплексной переменной. В качестве примера можно указать, скажем, фигурировавшую в § 6.3 функцию $y=\frac{1}{1+x^2}$ (см. ее график на рис. 6.3.3). Эта функция может быть разложена в ряд Маклорена (см. формулу (6.3.8)), который, однако, неожиданным образом сходится лишь при $|x|<1$, хотя никаких особенностей эта функция не имеет. Связано это с тем, что (аналитическая) функция комплексной переменной $w=\frac{1}{1+z^2}$ обращается в бесконечность в точках $z=\pm i$, расположенных на окружности $|z|=1$ — эти-то особенности и не позволяют продолжить ряд Маклорена за пределы круга $|z|<1$. Аналогичное происхождение имеет и неаналитичность в точке $x=0$ рассмотренной в § 1 функции $y=e^{-1/x^2}$ (ее график изображен на рис. 1.1). Для функции $w=e^{-1/z^2}$ комплексной переменной значение $z=0$ является, как говорят, *существенно особой точкой*: в этой точке функции нельзя приписать никакого значения. В самом деле, когда $z\to 0$ вдоль вещественной оси, т. е. $z=x$, где (вещественное) число x мало по абсолютной величине, то $-\frac{1}{z^2}=-\frac{1}{x^2}$ является очень большим по модулю отрицательным числом и $e^{-1/z^2}=e^{-1/x^2}$ очень мало. Но если $z\to 0$ вдоль мнимой оси плоскости комплексной переменной, т. е. $z=iy$, где y очень мало по абсолютной величине, то $-\frac{1}{z^2}=-\frac{1}{(iy)^2}=\frac{1}{y^2}$ будет очень большим положительным числом и, значит, $e^{-1/z^2}=e^{1/y^2}\to\infty$ при $y\to 0$. При стремлении $z\to 0$ в другом направлении величина e^{-1/z^2} может неограниченно приближаться к любому значению. Поэтому наша функция в точке $z=0$ никак не может быть аналитической (ей даже нельзя приписать никакого значения в этой точке!) — и ясно, что в окрестности этой точки она неразложима в степенной ряд (однако в окрестности любой другой точки она в степенной ряд разлагается).

Заметим, что, например, функция $w=f(z)$, где $w=0$ при $x<0$ и $w=1$ при $x>0$, не будет аналитической, причем дело здесь заключается не столько в том, что w имеет скачок, сколько в том, что в описании w величины x и y, т. е. вещественная и мнимая части комплексной переменной z, в некотором смысле неравноправны. По аналогичной причине не является аналитической и функция $w=z^*=x-iy$: ведь при составлении z^* мы по-разному поступаем с вещественной частью z (оставляем ее неизменной) и с мнимой ее частью (ее знак мы изменяем).

С другой стороны — и это обстоятельство является весьма принципиальным, — если в определение функции входит именно z, а не x и y в отдельности, не z^* и не $|z|$, то функция аналитична даже если ее нельзя выразить в виде конечной комбинации элементарных функций. Так, например, аналитической является функция, заданная рядом, скажем,

$$w(z)=\sum_n \frac{z^{3n}}{(3n)!}, \qquad (15.2.8)$$

или определенным интегралом, как, например,

$$w(z)=\int_0^z e^{-p^2}dp \qquad (15.2.9)$$

(о смысле записи (9), где z и p — комплексные числа, мы еще скажем ниже), или решением дифференциального уравнения, например

$$\frac{dw}{dz}=z^3-w^3, \quad \text{где } w=0 \text{ при } z=0,$$

— хотя во всех этих примерах w не выражается через z с помощью элементарных

[1] Ясно, что это не относится к искусственно образованным выражениям, в запись которых звездочка (*) входит лишь формально (простейший пример: $w=z^{**}$, т. е. $=(z^*)^*=z$).

функций. Но пример формул Коши—Римана (или условий гармоничности функций $u(x, y)$ и $v(x, y)$) показывает, что об аналитической функции можно высказать важные утверждения, не зная ее конкретного вида, а зная только что она является аналитической.

Более того, в общей теории функций комплексной переменной доказывается, что к а ж д а я аналитическая функция может быть выражена в виде р я д а или в виде и н т е г р а л а. Если для произвольной (вещественной) функции из наличия у нее первой производной вовсе еще не следует также и существование производных более высоких порядков, то для аналитических функций комплексной переменной (а в области комплексной переменной иные функции вообще не заслуживают внимания) дело обстоит совсем не так. Здесь из наличия у функции $w=f(z)$ первой производной $w'=f'(z)$ автоматически следует существование второй производной $f''(z)=\frac{dw'(z)}{dz}$ и вообще производных всех порядков (см. упр. 2). При этом в окрестности точки $z=z_0$, в которой функция $w=f(z)$ имеет производную, можно записать

$$f(z)=f(z_0)+\frac{f'(z_0)}{1!}(z-z_0)+\frac{f''(z_0)}{2!}(z-z_0)^2+\frac{f'''(z_0)}{3!}(z-z_0)^3+\ldots \tag{15.2.10}$$

А отсюда следует, что, зная функцию в с к о л ь у г о д н о м а л о й окрестности точки z_0 (этого достаточно для нахождения всех производных функции в точке z_0, т. е. для нахождения всех коэффициентов разложения (10)), мы можем «продолжить» функцию $w=f(z)$ за пределы этой окрестности, определив ее рядом (10). Далее, найдя таким образом значение функции $w=f(z)$ в точке $z=z_1$ и в соседних с z_1 точках, можно выписать аналогичное (10) разложение для точки $z=z_1$ и с помощью его найти значения w в новых точках комплексной плоскости и т. д. Возможность разложения аналитических функций комплексной переменной в ряд Тейлора (10) и оправдывает употребление одного и того же термина «аналитический» для функций вещественной и комплексной переменных. Она же указывает на то, что понятие «разумной», т. е. аналитической, функции комплексной переменной достаточно далеко от общего понятия «произвольной» функции, для полного задания которой необходимо указать все ее значения — значения во в с е х точках области определения функции.

Свойства аналитических функций комплексной переменной широко используются в теоретической физике. Законченной теории элементарных частиц в настоящее время еще нет. Однако достаточно предположить, что такая теория существует, что в принципе существуют какие-то неизвестные нам формулы, выражающие основные соотношения этой теории, чтобы отсюда уже можно было сделать содержательные выводы. Естественно полагать, что эти формулы, функциональные зависимости *аналитичны*. Тогда, даже не зная формул, можно кое-что высказать об элементарных частицах. В частности, можно утверждать, что масса частиц и античастиц одинакова. О нестабильных частицах можно утверждать, что их время жизни и время жизни их античастиц одинаково, хотя состав продуктов распада может быть разным. Объяснить здесь, как получаются такие далеко идущие выводы, совершенно невозможно. Однако хочется подчеркнуть, что предложение об аналитичности неизвестных нам формул играет здесь основную роль.

Первые попытки применения операций дифференцирования и интегрирования к комплексным величинам принадлежат создателям дифференциального и интегрального исчисления Готфриду Вильгельму Л е й б н и ц у и Иоганну Б е р н у л л и. В частности, простую формулу $\int\frac{a\,dx}{b^2+x^2}=\frac{a}{b}\operatorname{arctg}\frac{x}{b}$ Бернулли нашел в 1712 г. с помощью (совершенно правильных, но в то время еще недостаточно обоснованных) манипуляций, связанных с дифференцированием и интегрированием комплекснозначных выражений. В той же работе Бернулли получил в процессе своих преобразований красивую формулу $\left(\frac{\operatorname{tg}\alpha-i}{\operatorname{tg}\alpha+i}\right)^n=\frac{\operatorname{tg}n\alpha-i}{\operatorname{tg}n\alpha+i}$, родственную (тогда еще неизвестной) формуле Муавра (14.1.6). Однако полной ясности в этих вопросах ни у Лейбница, ни у Бернулли не было, о чем свидетельствует связанная с темой о комплексных функциях и их анализе дискуссия между ними о смысле логарифмов отрицательных чисел: Лейбниц считал эти логарифмы мнимыми, а Бернулли — вещественными. Публикация в 1745 г. переписки Лейб-

ница и И. Бернулли сразу же привлекла внимание крупнейших математиков того времени — Л. Эйлера и Ж. Л. Д'Аламбера (причем Эйлер стал здесь на сторону Лейбница, а Д'Аламбер — на сторону Бернулли). В частности, статья Эйлера «О споре между Бернулли и Лейбницем по поводу логарифмов отрицательных и мнимых чисел» (1749 г.) содержала современную теорию логарифмов комплексных чисел, изложенную в § 14.3, причем Эйлер особенно подчеркивал *многозначность* логарифмической функции, о чем Лейбниц и Бернулли не подозревали. Для Эйлера характерно было также специальное внимание к аналитическим функциям вещественной переменной, которые он называл «непрерывными» [2]; Эйлер даже склонен был считать, что это — единственные функции, заслуживающие внимания математиков. Работы Эйлера изобиловали конкретными результатами, сегодня относящимися к учению об аналитических функциях комплексной переменной, включая многочисленные образцы разложений таких функций в ряды и примеры использования подобных рядов. Впрочем, особенно глубоко в изучении свойств аналитических функций — как вещественных, так и комплексных — Эйлер не продвинулся. Так, например, он считал, что, зная значения вещественной аналитической функции $y=f(x)$, отвечающие любому (сколь угодно малому!) интервалу изменения независимой переменной x, можно восстановить значения функции на всей числовой прямой, а это, разумеется, неверно.

Д'Аламбер сделал принципиальный шаг вперед в области теории функций, четко оговорив, что и аргумент и значения функции могут быть как вещественными, так и комплексными — в этом пункте Эйлер в своих исследованиях следовал за Д'Аламбером. Аналитическую (записанную формулой) функцию $w=f(z)$, где $z=x+iy$, Д'Аламбер записывал как $u(x, y)+iv(x, y)$; в частности, путем дифференцирования функции z^k, где и z, и k — комплексные числа, причем k — постоянно, а z — переменно, он находит правильное выражение (в виде $u+iv$) для комплексной степени комплексного числа. Им же были впервые получены уравнения Коши—Римана (6): они содержатся в работе Д'Аламбера от 1752 г., посвященной плоскому движению идеальной (несжимаемой) жидкости. (О связи этой темы из механики с теорией аналитических функций комплексной переменной мы еще скажем ниже (см. § 17.3).) В 1761 г. Д'Аламбер вывел из уравнений (6) и равенства (7). Все эти работы по механике жидкостей и по теории аналитических функций комплексной переменной особенно активно продолжал Жозеф Луи Лагранж.

Подхватив исследования Д'Аламбера по гидромеханике, Эйлер в 1755 г. также (причем независимо от Д'Аламбера) пришел к уравнениям (6), которым он придавал большое значение и из которых получал целый ряд следствий (включая следствия геометрического характера, связанные с так называемыми *конформными преобразованиями* [3]). В 1776—1777 гг. Эйлер (ему было в тот период около 70 лет) начал пионерские работы по изучению (и приложению) интегралов от функций комплексной переменной (см. § 17.2). Однако отчетливое определение интеграла от (аналитической) функции комплексной переменной и представление об основных свойствах таких интегралов было впервые дано только Карлом Фридрихом Гауссом в 1811 г. в письме к известному немецкому математику и астроному Ф. В. Бесселю. Впрочем, опубликовано это письмо Гаусса было лишь в 1880 г., когда все содержавшиеся в нем результаты были уже известны из исследований Огюстена Коши, занимавшегося соответствующей темой начиная с 1813 г.

В некоторых учебниках теории функций комплексной переменной уравнения (6) называют «уравнениями Д'Аламбера—Эйлера». Нам, однако, кажется более справедливым традиционное наименование: «уравнения Коши—Римана». Дело в том, что ни Д'Аламбер, ни Эйлер теории функций комплексной переменной не создали: так, у них отсутствовали даже полноценные определения понятий производной и интеграла от аналитической функции и не была выяснена связь условий (6)

[2] В настоящее время этот термин имеет совершенно другой смысл (ср. с употреблением его Риманом (с. 433)).

[3] Так называются преобразования (плоскости или пространства), сохраняющие углы между линиями и в окрестности каждой точки имеющие характер преобразования подобия. Нетрудно видеть, что отображение $z \to w(z)$ плоскости комплексной переменной на себя, задаваемое аналитической функцией $w=f(z)$ с производной $f'(z)$, имеет именно такой характер: из того, что в окрестности точки $z=z_0$ выполняется приближенное равенство $w-w_0 \approx f'(z_0)(z-z_0)$ (где $w_0=f(z_0)$), и из правил действий над комплексными числами (см. гл. 14) следует, что вблизи z_0 это отображение складывается из сдвига плоскости в точку w_0 (переноса на вектор $\overrightarrow{z_0w_0}$), растяжения в ρ раз, где $\rho=|f'(z_0)|$, и поворота на угол φ, где $\varphi=\operatorname{Arg} f'(z_0)$.

с самим фактом существования производной $\frac{dw}{dz}=\lim\limits_{|\Delta z|\to 0}\frac{w(z+\Delta z)-w(z)}{\Delta z}$ функции комплексной переменной, не зависящей от направления смещения Δz. Последовательное изложение учения о (дифференцируемых, т. е. аналитических) функциях комплексной переменной дал впервые Коши в целом ряде обширных статей, важнейшей из которых была большая статья от 1825 г., посвященная «комплексному интегрированию». Может быть, еще большее значение в развитии теории аналитических функций (комплексной переменной) сыграла замечательная докторская диссертация Бернхарда Римана (1851 г.) «Основы общей теории функций одной комплексной переменной», явившаяся первым в истории науки связным изложением всей новой теории (из этой диссертации вырос поразительный по богатству материала и глубине идей курс лекций, прочитанный Риманом в Гёттингене зимой 1855/56 г. и летом 1856 г.[4]) и заложившая основы того, что сегодня называют «геометрической теорией аналитических функций». И настоящую главу мы закончим обширной цитатой из этой диссертации Римана,[5] которая охватывает основные моменты всего сказанного нами выше и выразительно демонстрирует, «как же все это началось»:

«Пусть через z обозначена переменная величина, которая, изменяясь неограниченно, может принимать всевозможные действительные значения; если каждому ее значению соответствует единственное значение некоторой величины w, то w называют функцией z; если, кроме того, в то время как переменная z, изменяясь непрерывно, пробегает все значения, заключенные между двумя постоянными пределами, величина w также изменяется непрерывно, то внутри указанного промежутка эта функция называется непрерывной. Ясно, что такое определение никоим образом не устанавливает связи между отдельными значениями функции: если поведение функции указано лишь в некотором определенном промежутке, то за пределы этого промежутка ее можно продолжить совершенно произвольно.

Зависимость величины w от величины z может быть задана посредством математической формулы так, что значение w, соответствующее каждому значению z, получается в результате определенных математических операций, производимых над числовым значением z. . .

Пусть область изменения величины z не ограничивается действительными значениями, а распространяется также и на комплексные значения вида $x+yi$, где $i=\sqrt{-1}$. Пусть $x+yi$ и $(x+dx)+(y+dy)i$ — два бесконечно мало различающихся между собой значения переменной z, которым соответствуют значения $u+vi$ и $(u+du)+(v+dv)i$ переменной w. В таком случае, если зависимость переменной w от переменной z считать совершенно произвольной, то отношение $\frac{du+dvi}{dx+dyi}$, вообще говоря, изменяется вместе с изменением dx и dy; если положить $dx+dyi=\varepsilon e^{\varphi i}$, то ясно видно, что

$$\frac{du+dvi}{dx+dyi}=\frac{1}{2}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1}{2}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)i+$$
$$+\frac{1}{2}\left[\left(\frac{\partial u}{\partial x}-\frac{\partial v}{\partial y}\right)+\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right)i\right]\frac{dx-dyi}{dx+dyi}=$$
$$=\frac{1}{2}\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\frac{1}{2}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)i+$$
$$+\frac{1}{2}\left[\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+\left(\frac{\partial v}{\partial x}+\frac{\partial u}{\partial y}\right)i\right]e^{-2\varphi i}.$$

Однако всякий раз, когда функция w переменной z определена посредством формулы, содержащей простые математические операции, оказывается, что значение производной $\frac{dw}{dz}$ не зависит от выбора значения дифференциала[6] dz. Таким образом, становится ясным, что указанным способом можно выразить не всякую зависимость комплексной переменной w от комплексной переменной z.

Отмеченное нами характерное свойство всех тех функций, которые могут быть определены с помощью элементарных математических операций, мы положим в основу нашего дальнейшего исследования . . . именно, мы будем исходить из следующего определения . . .:

некоторая комплексная переменная w называется функцией другой комплексной переменной z, если она меняется вместе с ней та-

---

[4] Этот курс лекций, сыгравший огромную роль в истории науки, слушали всего три человека (впрочем, на с. 182 мы упоминали об оказавшемся весьма важным лекционном курсе И. Бернулли, имевшем всего одного слушателя).

[5] См.: *Риман Б.* Сочинение. М.; Л.: Гостехиздат, 1948, с. 49—87.

[6] Это утверждение, очевидно, выполнено во всех тех случаях, когда выражение $\frac{dw}{dz}$ можно получить из выражения w через z путем обычных правил дифференцирования. . . (Примеч. Римана. — *Я. З.*, *И. Я.*).

ким образом, что значение производной $\frac{dw}{dz}$ не зависит от значения дифференциала dz.

... Если отношение $\frac{du+dvi}{dx+dyi}$ мы представим в виде

$$\frac{\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial x}\,i\right)dx+\left(\frac{\partial v}{\partial y}-\frac{\partial u}{\partial y}\,i\right)dyi}{dx+dyi},$$

то становится ясным, что оно в том и только в том случае имеет одно и то же значение для двух любых значений dx и dy, если имеют место равенства

$$\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y} \quad \text{и} \quad \frac{\partial v}{\partial x}=-\frac{\partial u}{\partial y}.$$

Итак, эти условия являются достаточными и необходимыми для того, чтобы $w=u+vi$ являлось функцией $z=x+yi$. Отсюда для отдельных членов этой функции вытекают условия

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0, \quad \frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}=0,$$

которые служат основой при изучении свойств, принадлежащих каждому члену таких функций, рассматриваемому вне зависимости от другого члена».

Упражнения

15.2.1. Проверьте выполнимость условий Коши—Римана для функций комплексной переменной, фигурирующих в упр. 14.1.1, и для функции $w=\ln z$.

15.2.2. Пусть $w=f(z)=u(x, y)+iv(x, y)$ — аналитическая функция комплексной переменной, а $f'(z)=\frac{\partial w}{\partial z}=u_1(x, y)+iv_1(x, y)$ — ее производная; докажите, что из выполнимости условий Коши—Римана для функции $f(z)$ автоматически вытекает и выполнимость их для функции $f'(z)$.

Глава 16
ЗАМЕЧАТЕЛЬНАЯ ДЕЛЬТА-ФУНКЦИЯ ДИРАКА

§ 1. Различные способы определения функции

Те функции, которые мы изучали до сих пор, обычно задавались с помощью формул. Это значит, что задавался способ вычисления значений функции при любом значении независимой переменной. Такое задание можно назвать *алгоритмическим* (алгоритм — способ вычисления). Ведь формула $y=f(x)=2x+3x^2$ означает следующее: чтобы найти значение y при данном x, возьми x, умножь на 2; возведи x в квадрат, умножь на 3; затем сложи два полученных числа. Иначе определялись тригонометрические функции — с помощью геометрических понятий, измерения дуг и отрезков в круге; однако и здесь можно говорить об алгоритме вычисления их значений, состоящем в обращении к *тригонометрическим таблицам* (или к микрокомпьютеру с соответствующими кнопками).

Нашей задачей раньше было изучение свойств функций, заданных таким образом: исследование закона их возрастания и убывания, нахождение максимумов и минимумов и т. д. Изучение этих свойств приводит к новым способам определения функций. Так, например, функцию $f=e^x$ можно определить как такую функцию, производная которой равна самой функции: $\frac{df}{dx}=f$ при дополнительном условии $f(0)=1$. Синус, т. е. функция $\varphi=\sin x$, и косинус, т. е. $\psi=\cos x$, можно определить как функции, удовлетворяющие одному и тому же уравнению

$$\frac{d^2\varphi}{dx^2}=-\varphi, \quad \frac{d^2\psi}{dx^2}=-\psi,$$

но различным начальным условиям:

$$\varphi(0)=0, \quad \left.\frac{d\varphi}{dx}\right|_0=1;$$

$$\psi(0)=1, \quad \left.\frac{d\psi}{dx}\right|_0=0.$$

Эти определения оказываются во многих отношениях более «по существу», более тесно связанными с применениями показательной и тригонометрических функций ко многим задачам физики, например к задачам колебаний, по сравнению с определениями

$$e^x=\left[\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n\right]^x$$

$$\left(\text{или} \quad e^x=\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n\right)$$

и определениями синуса и косинуса как некоторых отрезков в круге или как отношений сторон прямоугольного треугольника.

Любопытно, что определения e^x, $\sin x$, $\cos x$ с помощью дифференциальных уравнений оказываются удобными для электронных вычислительных машин. Если по ходу расчета необходимо подставлять в формулу значения e^x при различных x, то нужно взять таблицу и выписывать из нее нужные значения e^x. При работе на машине оказывается легче и быстрее не обращаться к таблице, а вычислять на машине значения функции e^x шаг за шагом по приближенной формуле $e^{x+\Delta x}\approx e^x(1+\Delta x)$, вытекающей из дифференциального уравнения $\frac{dy}{dx}=y$ для функции e^x (или по более точным формулам, основанным на том же уравнении). То же относится и к функциям $\sin x$ и $\cos x$; их удобно вычислять одновременно [1]:

$$\sin(x+\Delta x)\approx\sin x+\Delta x\cos x,$$

$$\cos(x+\Delta x)\approx\cos x-\Delta x\sin x.$$

Итак, один общий подход к понятию функции заключается в задании формулы как способа ее вычисления и последующем исследовании такой функции. Возможен и другой подход: можно искать функцию с определенными общими свойствами, с тем чтобы потом на основании этих свойств попытаться найти ту формулу, которая описывает интересующую нас функцию. Такая про-

[1] Продумайте, откуда получаются эти формулы, и проверьте, согласуются ли они с уравнением для второй производной синуса и косинуса.



ПОЛЬ ДИРАК

цедура всегда применяется при обработке экспериментальных данных и при подборе эмпирических (т. е. основанных на опыте) формул для описания того или иного явления. Здесь же наша цель — построить таким способом одну замечательную и весьма необычную функцию, полезную и важную как для самой математики, так и для ее применений.

§ 2. Дирак и его функция

Поль Адриан Морис Д и р а к — крупнейший английский физик-теоретик — прославился в 1929 г. Он строил такую теорию, которая описывала бы движение электронов в электрических и магнитных полях с любой скоростью, вплоть до скорости, близкой к скорости света. Речь идет о квантовой теории, которая объясняет также тот факт, что в атоме электроны могут двигаться только по определенным орбитам, с определенными значениями энергии. Дирак знал, что электрон обладает некоторым моментом вращения, т. е. подобен вращающемуся волчку, и учитывал это при создании своей теории. Когда же теория была им создана, оказалось, что из нее следует еще и вывод, не предусмотренный Дираком: возможно существование частиц с такой же массой, как электроны, но с противоположным (положительным) знаком заряда. В течение двух лет считали, что теория Дирака хороша для описания движения электрона, а вывод о частицах с положительным зарядом ошибочен — и когда от него удастся избавиться, теория станет совсем отличной.

Но в 1932 г. частицы с положительным зарядом — *позитроны*, или, как их еще называют, «античастицы электрона», были открыты. Их появление в теории из недостатка превратилось в триумф, в главное достижение теории: открытие Дирака было первым примером появления новых частиц «на острие пера теоретика». Этот пример поучителен с точки зрения взаимоотношений опыта и теории: теория основывается на определенных данных опыта, но последовательное логическое и математическое развитие теории выводит за пределы того материала, который был положен в ее основу, приводит к новым предсказаниям.

Дирак является не только одним из лучших физиков-теоретиков нашей планеты, но и замечательным математиком. В своем классическом труде «Принципы квантовой механики» Дирак ввел и широко использовал новую функцию, которой он дал обозначение $\delta(x)$ (читается: *дельта-функция* или *дельта-функция Дирака*).

Дельта-функцию можно описать следующим образом: $\delta(x) \equiv 0$ при любом $x \neq 0$ (т. е. при $x < 0$, и при $x > 0$) и $\delta(0) = \infty$. Кроме того, функция $\delta(x)$ должна удовлетворять условию

$$\int_{-\infty}^{+\infty} \delta(x)\,dx = 1. \qquad (16.2.1)$$

Наглядно можно представить себе график функции, похожей на $\delta(x)$; он изображен на рис. 1. Чем более узкой мы сделаем полоску между левой и правой ветвями, тем выше должна быть эта полоска, для того чтобы площадь полоски (т. е. интеграл) сохраняла заданное значение, равное 1. При сужении полоски (и одновременном росте графика) мы приближаемся к выполнению условия $\delta(x)=0$ при $x \neq 0$, т. е. изображенная на графике функция приближается к дельта-функции. В одном из следующих параграфов эти соображения будут использованы для построения формул, задающих дельта-функцию. Здесь же мы продолжим исследования ее общих свойств.

Следует, конечно, сразу же подчеркнуть, что определяющие дельта-функцию «равенства»

$$\delta(x)=\begin{cases}0 & \text{при } x\neq 0,\\ \infty & \text{при } x=0,\end{cases}\qquad \int_{-\infty}^{+\infty}\delta(x)\,dx=1 \tag{16.2.2}$$

не следует понимать слишком уж буквально: ведь с точки зрения классической (привычной нам) математики эти условия не имеют смысла, противоречивы. Разумеется, описанный выше процесс постепенного вытягивания (и сжатия) «полосок» поясняет смысл дельта-функции; наивное ее «определение» (2) призвано лишь стимулировать наше воображение, напоминая об этом процессе.

Ясно, что, определяя дельта-функцию, мы вводим в рассмотрение совсем новый объект, не похожий ни на один из тех, с которыми нам ранее приходилось встречаться (вот уж когда нашу функцию никак нельзя назвать «аналитической» (см. гл. 15)!). Однако описание этого нового объекта не выглядит сложно и оно не должно нас смущать. Ведь, скажем, и число $\sqrt{2}$, с которым математикам столь часто приходится иметь дело, с точки зрения арифметики настоящим числом не является: оно задается не точным своим значением, а лишь совокупностью приближений 1; 1,4; 1,41; 1,414; 1,4142;. . . к этому числу (которые, впрочем, характеризуют $\sqrt{2}$ достаточно полно для того, чтобы мы могли свободно этим числом пользоваться). Аналогично и функция $\delta(x)$ задается не точными своими значениями, а лишь приближениями к ней — высокими и узкими «ступеньками» [1]; эти приближения полностью нашу функцию характеризуют.

Дирак и другие физики многие годы свободно и продуктивно пользовались дельта-функциями, не заботясь о строгом определении этих удивительных объектов, подобно тому как математики много веков использовали иррациональные числа до того, как в конце XIX в. были созданы первые теории вещественного числа. Математическое обоснование удивительных функций, подобных функции $\delta(x)$ («обобщенных функций», как их обычно в настоящее время называют), дали академик С. Л. Соболев, член-корреспондент Академии наук И. М. Гельфанд, француз Л. Шварц и другие ученые; ему посвящены многие книги.[2]

Возможно, лучше, чем (2), объясняет смысл функции Дирака следующее важнейшее равенство:

$$\int_{-\infty}^{+\infty} f(x)\,\delta(x)\,dx=f(0). \tag{16.2.3}$$

В самом деле, так как $\delta(x)=0$ при $x\neq 0$, то значение стоящего в левой части (3) интеграла не может зависеть

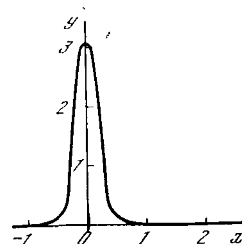


Рис. 16.2.1

от значений $f(x)$ ни при каком $x\neq 0$. Существенно только значение $f(x)$ там, где $\delta(x)\neq 0$, т. е. при $x=0$. В той узкой области, где $\delta(x)\neq 0$ (в пределе ширина этой области стремится к 0 (см. рис. 1)), $\delta(x)$ умножается на $f(0)$. Следовательно, из условий (2) вытекает формула (3).

Все эти рассуждения можно провести и в обратном порядке, т. е. можно сказать, что $\delta(x)$ есть такая функция, для которой при любом виде вспомогательной функции $f(x)$ имеет место формула (3). Это одно условие приводит ко всем заключениям о виде $\delta(x)$, которые были раньше использованы для ее определения. Из формулы (3)

---

[1] Для желающих мы готовы выписать тут «десятичные приближения» δ-функции: $y=0$ при $|x|>1$, $y=0{,}5$ при $|x|<1$; $y=0$ при $|x|>0{,}1$, $y=5$ при $|x|<0{,}1$; $y=0$ при $|x|>0{,}01$, $y=50$ при $x<0{,}01$; . . . (удобнее еще «сгладить» эти ступеньки, чтобы все функции были непрерывными); ниже мы укажем другие еще более удобные приближения δ-функции.

[2] Пожалуй, простейшими из них являются две небольшие брошюры польских математиков: *Микусинский Я.*, *Сикорский Р.* Элементарная теория обобщенных функций. М.: Изд-во иностр. лит., 1959, вып. 1; 1963, вып. 2. В частности, в вып. 1 этой книги авторы вводят обобщенные функции как пределы последовательностей обыкновенных функций, скажем функций $\varphi_n(x)$, где $\varphi_n(x)=0$ при $|x|>\frac{1}{10^n}$; $\varphi_n(x)=\frac{10^n}{2}$ при $|x|<\frac{1}{10^n}$; см. сноску 1. (Разумеется, понятие «предел функции» нуждается в строгом определении.)

следует, что $\delta(x)=0$ при $x\neq 0$, что[3] $\int \delta(x)\,dx=1$ и что $\delta(0)=\infty$.

Укажем в заключение, что, конечно, и понимание соотношения (3), подобно определению (2) дельта-функции, также требует известного усилия мысли, поскольку под знаком интеграла в правой части (3) стоит нечто, что лишь с натяжкой можно назвать «функцией». Однако при «грубой» работе с дельта-функцией как сомнительное ее определение (2), так и равенство (3) оказываются достаточными для применений этого нового понятия; их следует всегда иметь в виду.

Приведем еще несколько очевидных следствий из определения $\delta(x)$. По общему правилу замены переменных (особенно подробно оно было изложено в § 1.7) функция $\delta(x-a)$ смещена на a единиц вправо по сравнению к $\delta(x)$, т. е. $\delta(x-a)=\infty$ при $x=a$ и $\delta(x-a)=0$ при $x\neq a$. Соответственно этому

$$\int_{-\infty}^{+\infty} f(x)\,\delta(x-a)\,dx=f(a). \qquad (16.2.3а)$$

Далее, легко убедиться, рассматривая кривую того вида, который изображен на рис. 1, что график функции $b\delta(x)$ в b раз выше, чем график $\delta(x)$, а график $\delta(cx)$ в $|c|$ раз уже, чем график $\delta(x)$, так что площадь под кривой $\delta(cx)$ в $|c|$ раз меньше, чем площадь под графиком $\delta(x)$ (разумеется, о всех этих «графиках» можно говорить лишь условно). Поэтому

$$\int f(x)\,b\delta(x)\,dx=bf(0), \qquad (16.2.4)$$

$$\int f(x)\,\delta(cx)\,dx=\frac{1}{|c|}f(0). \qquad (16.2.5)$$

Сравнение формул (4) и (5) позволяет сказать, что $\delta(cx)=\frac{1}{|c|}\delta(x)$.

Формула (4) совершенно очевидна, но можно надеяться, что для читателя, прошедшего через испытания всеми предыдущими главами книги, будет сразу понятна и формула (5). Ее легко получить формально заменой переменных: $u=|c|\,x$, $dx=\frac{1}{|c|}du$. Здесь мы используем еще тот факт, что определяемую согласно (2) функцию $\delta(x)$ можно считать **четной**: $\delta(-x)=\delta(x)$.

Упражнения

16.2.1. Покажите, что если функция $\varphi(x)$ имеет единственный нуль x_0, так что $\varphi(x_0)=0$ и $\varphi(x)\neq 0$ при $x\neq x_0$, то $\delta(\varphi(x))=\frac{1}{|\varphi'(x_0)|}\delta(x-x_0)$. Что представляет собой функция $\delta(\psi(x))$, если функция $\psi(x)$ обращается в 0 при нескольких значениях x?

16.2.2. Чему равен $\int_{-\infty}^{+\infty}\psi(x)\,\delta(\sin x)\,dx$?

§ 3. Разрывные функции и их производные

Исследуем интеграл от функции $\delta(x)$, рассматриваемый как функция верхнего предела интегрирования, т. е. функцию

$$\theta(x)=\int_{-\infty}^{x}\delta(z)\,dz. \qquad (16.3.1)$$

Легко убедиться, что график этой функции имеет вид «ступеньки» (рис. 1). В самом деле, пока $x<0$, область интегрирования в формуле (1) такова, что в ее пределах всюду $\delta(x)=0$; поэтому при $x<0$ и $\theta(x)=0$. Если же $x>0$, то в область интегрирования включается окрестность начала координат, где $\delta(0)=\infty$. С другой стороны, так как при $x>0$ также $\delta(x)=0$, то значение интеграла не изменяется, когда верхний предел меняется от 0,1 или даже от 0,000001 до 1, или до 10, или до ∞. Следовательно, при **любом** $x>0$

$$\theta(x)=\int_{-\infty}^{x}\delta(z)\,dz=\int_{-\infty}^{+\infty}\delta(z)\,dz=1,$$

как и показано на рис. 1.

Таким образом, с помощью дельта-функции сконструирована простейшая разрывная функция $\theta(x)$, такая, что $\theta(x)=0$ при $x<0$ и $\theta(x)=1$ в области $x>0$[1]. Эти простые соображения

[3] Интеграл без обозначения пределов интегрирования здесь будет всегда пониматься распространенным по полному интервалу изменения переменной интегрирования: от $-\infty$ до $+\infty$.

[1] В самой точке $x=0$ функция $\theta(x)$ «совершает прыжок»; конкретизировать ее значение в этой точке не следует, достаточно сказать, что здесь значение функции переходит от $\theta=0$ к $\theta=1$.

позволяют более последовательно, без кажущихся исключений и многословных оговорок, подойти к вопросу о производной от функции, имеющей разрывы.

Не зная дельта-функции, приходится говорить, что производные нельзя находить там, где функция разрывна. Но вот мы построили разрывную функцию $\theta(x)$. Общее правило о связи

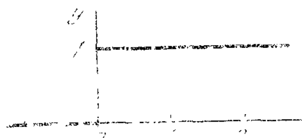


Рис. 16.3.1

между интегралом и производной (см. § 3.3) имеет вид:

$$\text{если } F(x) = \int_{x_0}^{x} y(z)\,dz, \text{ то } y(x) = \frac{dF}{dx}.$$

Применяя его к (1), получим

$$\frac{d\theta(x)}{dx} = \delta(x). \qquad (16.3.2)$$

Значит, для производной разрывной функции не надо делать исключений: просто производная такой функции равна «особой» функции — дельта-функции.

Научившись справляться с производной простейшей разрывной функции, можно легко находить производные и в более сложных случаях. Приведем несколько примеров. Пусть

$$y = x \text{ при } x < 1, \quad y = x - 2 \text{ при } x > 1. \qquad (16.3.3)$$

График функции $y(x)$ изображен на рис. 2. Разрыв имеет место при $x=1$. Величина разрыва $y(1+0)-y(1-0)=-2$. (Мы пользуемся здесь записью $y(1+0)$, понимая под этим предельное значение y при приближении x к 1 справа (со стороны $x > 1$); запись $y(1-0)$ означает то же, но здесь речь идет о приближении x к 1 слева (см. рис. 2).) В силу этого можно формально записать

$$\frac{dy}{dx} = 1 - 2\delta(x-1). \qquad (16.3.4)$$

Такая запись много содержательней и информативней утверждения о том, что $\frac{dy}{dx} = 1$ всюду, кроме точки $x=1$, где функция терпит разрыв и не имеет производной.

Дельта-функция — это типичное дитя XX в. Прошлому XIX в. свойственно было пристрастие облекать многие свои суждения — и правильные, и не совсем правильные, и совсем неправильные — в форму «невозможностей». Невозможно изобрести вечный двигатель, невозможны превращения химических элементов, невозможно изменение общей массы вещества, невозможно доказать реальное существование атомов, невозможно определить состав звезд, невозможно найти производную разрывной функции. Наш XX в. находит много конструктивных решений для того, что казалось невозможным в XIX столетии, скажем для превращений химических элементов или для уменьшения массы путем превращения ее в энергию; в частности, и дельта-функция решает вопрос о производной функции в точке ее разрыва (во всяком случае, в точке разрыва, имеющего вид конечного скачка). В самом деле, ведь запись (4) в одной строчке указывает и на сам факт разрыва функции $y(x)$ (раз в выражение для $\frac{dy}{dx}$ вошла дельта-функция), и характер разрыва (скачок), и место разрыва $x=1$, и величину разрыва (ее характеризует коэффициент -2 при дельта-функции).

Интегрируя (4) при начальном условии: $y=0$ при $x=0$, возможно полностью восстановить график $y(x)$. Правда, в случае функции (3) нам «повезло»; другими словами, был специально выбран простой случай, когда слева и

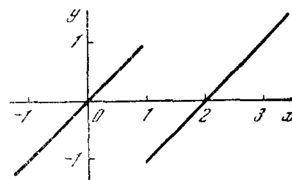


Рис. 16.3.2

справа производная выражается одной и той же формулой. Это, конечно, не обязательно — с чего бы производной быть непрерывной, если сама функция терпит разрыв!

Рассмотрим более сложный пример: $y=-x^2$ при $x < 1$; $y = x^2$ при $x > 1$ (рис. 3). Ясно, что $y'=-2x$ при $x < 1$ и $y'=2x$ при $x > 1$. Разрыву соответствует значение $y'=4\delta(x-1)$ производной. Теперь мы можем произвольно

присоединить точку $x=1$ к левой области определения функции и записать $y'=-2x+4\delta(x-1)$ при $x\leqslant 1$; $y'=2x$ при $x>1$, либо — другой вполне равноправный с первым вариант — можно присоединить $x=1$ к правой части области определения $y(x)$ и считать, что $y'=-2x$ при $x<1$; $y'=2x+4\delta(x-1)$ при $x\geqslant 1$. (Обратите внимание на то, как расставлены знаки $<$ (меньше) и $\leqslant$ (меньше или равно), $>$ (больше) и $\geqslant$ (больше или равно) в формулах, — нужно внимательно следить за тем, чтобы не записать дельта-функцию дважды [2].) Можно писать также

$$y'=\varphi(x)+4\delta(x-1),$$

где

$$\varphi(x)=\begin{cases}-2x \text{ при } x<1,\\ 2x \text{ при } x>1.\end{cases}$$

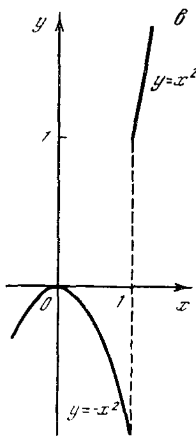


Рис. 16.3.3

Чтобы проверить запись, проинтегрируйте выражение производной и получите снова исходную (разрывную!) функцию.

Иногда в математике и ее приложениях оказывается полезной так называемая *знаковая функция* [3] $\operatorname{sgn}(x)$. Она определяется так: $\operatorname{sgn}(x)=-1$ при $x<0$ и $\operatorname{sgn}(x)=1$ при $x>0$. Можно записать $\operatorname{sgn}(x)=\frac{x}{|x|}$, где $|x|$ — модуль, или абсолютное значение, x. График $\operatorname{sgn}(x)$ показан на рис. 4, *а*. Легко убедиться, что

$$\operatorname{sgn}(x)=-1+2\theta(x)=-1+2\int_{-\infty}^{x}\delta(x)\,dx. \qquad (16.3.5)$$

[2] Таким образом, производная разрывной функции $f(x)$ в точке разрыва $x=x_0$ имеет вполне определенное «значение» $\delta(x-x_0)$, в то время как для самой функции $f(x)$ нет смысла спрашивать, чему она равна в точке $x=x_0$ (во всяком случае, этот вопрос является бессмысленным почти во всех прикладных задачах).

[3] sgn — сокращенное обозначение латинского слова signum — «знак» (запись $\operatorname{sgn}(x)$ часто читают как «сигнум икс»). (Иногда также условно считают, что $\operatorname{sgn}(0)=0$ — нам это физически мало осмысленное предположение не будет нужно.)

Рассмотрим далее саму функцию $|x|$; ее график изображен на рис. 4, *б*. Найдем производные этой функции:

$$\frac{d|x|}{dx}=-1 \text{ при } x<1,$$

$$\frac{d|x|}{dx}=1 \text{ при } x>1,$$

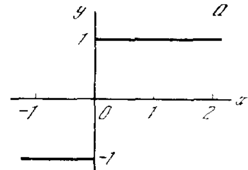


Рис. 16.3.4

или, кратко, с помощью новой функции, которую мы только что узнали:

$$\frac{d|x|}{dx}=\operatorname{sgn} x. \qquad (16.3.6)$$

Продифференцируем (6) еще раз, воспользовавшись соотношением (5):

$$\frac{d^2|x|}{dx^2}=2\delta(x). \qquad (16.3.7)$$

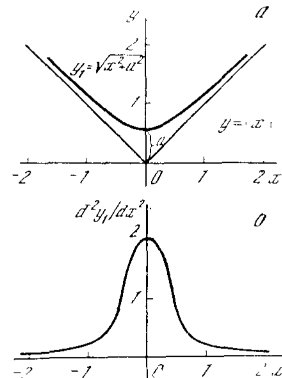


Рис. 16.3.5

Это очень важная формула, которую надо прочувствовать. Ведь мы знаем, что вторая производная функции связана с кривизной ее графика (см. § 7.9); для прямой линии кривизна равна нулю, ибо равна нулю вторая производная линейной функции. Казалось бы, что поскольку график $|x|$ состоит из двух прямых, на каждой из которых $\frac{d^2|x|}{dx^2}=0$, то можно так же просто сказать, что $\frac{d^2|x|}{dx^2}$ равно нулю везде.

Но так сказать нельзя — ясно, что дело тут заключается в изломе в точке $x=0$, где сходятся две прямые. Но как убедиться в том, что в точке излома вторая производная равна именно $2\delta(x)$? Для этого давайте скруглим излом: возьмем функцию

$$y_1=\sqrt{x^2+a^2}. \qquad (16.3.8)$$

Эта функция тем ближе к ломаной $y=|x|$, чем меньше a (рис. 5, *а*); ясно, что при $a=0$ формула (8) дает $y_1=|x|$. Функция (8) — гладкая, всюду дифференцируемая; по правилам гл. 4 легко найдем:

$$\frac{dy_1}{dx}=\frac{x}{\sqrt{x^2+a^2}},$$

$$\frac{d^2y_1}{dx^2}=\frac{a^2}{(x^2+a^2)^{3/2}}. \qquad (16.3.9)$$

На рис. 5, *б* представлен график функции $\frac{d^2y_1}{dx^2}$ (на рис. 5 принято $a=0{,}5$). Легко убедиться, что этот график становится все выше и у́же при уменьшении a, и при всех a интеграл $\int\frac{d^2y_1}{dx^2}dx=2$ (см. упр. 3), откуда в пределе при $a\to 0$ и получается формула $\frac{d^2|x|}{dx^2}=2\delta(x)$.

Упражнения

16.3.1. Нарисуйте графики (разрывных) функций и выпишите их (первые) производные: а) $y=x$ при $x<1$, $y=x-1$ при $x>1$; б) $y=\frac{e^{1/x}}{1+e^{1/x}}$.

16.3.2. Найдите кривизну линии (8) $y_1=y_1(x)$; рассмотрите, как меняется кривизна при $a\to 0$.

16.3.3. Докажите, что для функции (8) $\int_{-\infty}^{+\infty}y_1''(x)dx=2$.

§ 4. Представление дельта-функции формулами

В конце предыдущего параграфа невзначай было получено выражение для функции

$$y=\frac{a^2}{2(x^2+a^2)^{3/2}}, \qquad (16.4.1)$$

которая в пределе при $a\to 0$ приближается к $\delta(x)$[1]. Остановимся на вопросе об аналитическом выражении дельта-функции подробнее.

Пусть $\varphi(x)$ — произвольная функция; мы потребуем только обращения ее в нуль при $x=\pm\infty$ (т. е. потребуем, чтобы $\lim\varphi(x)=0$ при $x\to\pm\infty$) и отличия от нуля интеграла $I=\int\varphi(x)dx$. Умножая φ на соответствующую константу, всегда можно сделать интеграл I равным 1. Предположим, что мы так уже и поступили, т. е. что $\int\varphi(x)dx=1$. Ясно, что φ имеет максимум где-то между $-\infty$ и $+\infty$. Простейшие примеры представляют собой функции везде

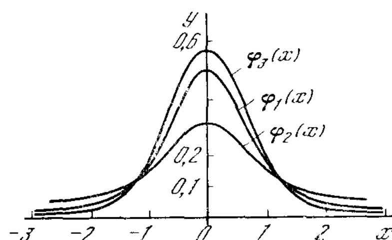


Рис. 16.4.1

положительные и четные, т. е. симметричные относительно оси y (см. § 1.7). Вот несколько конкретных примеров:

$$\varphi_1(x)=\frac{1}{2(1+x^2)^{3/2}},\quad \varphi_2(x)=\frac{1}{\pi}\frac{1}{1+x^2},$$

$$\varphi_3(x)=\frac{1}{\sqrt{\pi}}e^{-x^2}.$$

Графики этих функций имеют колоколообразный вид (рис. 1); если их привести к одинаковой высоте (см. ниже), то на глаз их даже не легко отличить. (Впрочем, последняя (экспоненциальная) функция при больших x (т. е. вдали от максимума) гораздо теснее приближается к оси абсцисс, чем остальные[2].) Теперь вспоминаем, как нужно поступать для того, чтобы увеличить в n раз высоту колокола и уменьшить в m раз его ширину: для этого образуем функцию $n\varphi(mx)$. Если мы хотим сохранить площадь под колоколом, то придется считать, что $n=m$. Итак, функция $n\varphi(nx)\to\delta(x)$ при $n\to\infty$, или, иначе,

[1] Отметьте коэффициент 2 в знаменателе правой части формулы (1): без этой двойки у нас получалось бы $2\delta(x)$.

[2] Пересечение на рис. 1 всех трех кривых φ_1, φ_2, φ_3 приблизительно в одних и тех же двух точках является, конечно, случайностью.

$\delta(x) = \lim\limits_{n\to\infty} n\varphi(nx)$. Легко проверить и формально, путем подстановки $z=nx$, что $\int n\varphi(nx)\,dx = \int \varphi(z)\,dz = \int \varphi(x)\,dx = 1$.

Таким образом, трем вариантам $\varphi(x)$ соответствуют следующие три варианта представления дельта-функции:

$$\delta(x) = \lim_{n\to\infty} \frac{n}{2(1+n^2x^2)^{3/2}},$$
$$\delta(x) = \lim_{n\to\infty} \frac{n}{\pi(1+n^2x^2)}, \qquad (16.4.2)$$
$$\delta(x) = \lim_{n\to\infty} \frac{n}{\sqrt{\pi}} e^{-n^2x^2}.$$

Проверим, что тот рецепт, который был предложен раньше:

$$\delta(x) = \lim_{a\to 0} \frac{a^2}{2(x^2+a^2)^{3/2}},$$

подходит под это определение; для этого перепишем

$$\frac{a^2}{2(x^2+a^2)^{3/2}} = \frac{a^2}{2a^3\left(\frac{x^2}{a^2}+1\right)^{3/2}} = \frac{1}{2a\left(\frac{x^2}{a^2}+1\right)^{3/2}}$$

и обозначим $n=\frac{1}{a}$; получим первый вариант представления (2).

Итак, нет одной определенной простой формулы, которая давала бы $\delta(x)$: ясно, что одних слов $\delta(0)=\infty$ мало; чтобы определить δ, надо еще суметь показать, что это как раз та бесконечность, которая нужна. Однако $\delta(x)$ может быть получена как результат предельного перехода $(n\to\infty)$ из вполне «добропорядочных», хорошо определенных функций x, в которые вспомогательная величина n входит как параметр. При этом особенно нужно подчеркнуть, что δ можно получить таким предельным переходом из р а з н ы х функций φ. Пока число n остается конечным, отвечающие разным $\varphi(x)$ функции $n\varphi(nx)$, разумеется, отличаются одна от другой, в частности,

$$I = \int f(x)\,n\varphi(nx)\,dx \neq f(0), \qquad (16.4.3)$$

— и только при $n\to\infty$ все функции $n\varphi(nx)$ стремятся к одному и тому же пределу $\delta(x)$ и все интегралы[3] (3) — к $f(0)$:

$$\lim_{n\to\infty} I = f(0). \qquad (16.4.4)$$

Произвол, имеющий место в выборе первоначальной $\varphi(x)$, из которой получается $\delta(x)$, полностью соответствует сути дела. В § 17.4 будут рассмотрены примеры применения $\delta(x)$ к физике. Описание какого-либо воздействия, т. е. какой-то конечной функции $\psi(x)$, с помощью дельта-функции возможно и целесообразно именно тогда, когда детальная форма воздействия, т. е. истинная зависимость его от x, несущественна, а важен лишь интеграл.

Приведенные выше примеры, конечно, вовсе не исчерпывают разнообразные функции $\varphi(x)$, из которых можно «изготовить» $\delta(x)$. Можно отказаться от симметрии $\varphi(x)$: при переходе к $n\varphi(nx)$ и увеличении n сокращается и расстояние максимума от $x=0$, т. е. даже асимметрическая функция приближается к $\delta(x)$. Вот пример этого: функция $e^{-(x-1)^2}/\sqrt{\pi}$ при нашей подстановке переходит в функцию $ne^{-(nx-1)^2}/\sqrt{\pi}$, с максимумом в точке $x=1/n$.

Можно также отказаться от записи $\varphi(x)$ простой единой формулой, обеспечивающей гладкость $\varphi(x)$. Так, можно взять саму функцию $\varphi(x)$ разрывной, скажем положить

$$\varphi(x) = \begin{cases} 1/2 & \text{при } -1<x<1, \\ 0 & \text{при } x<-1 \text{ и } x>1. \end{cases} \qquad (16.4.5)$$

Предельный переход заключается в том, что мы полагаем

$$\varphi_n(x) = \frac{n}{2} \text{ при } -1<nx<1, \text{ т. е. при } -1/n<x<1/n,$$
$$\varphi_n(x) = 0 \text{ при } x<-1/n \text{ и } x>1/n \qquad (16.4.6)$$

и устремляем $n\to\infty$. (Нарисуйте график функции $\varphi(x)$, задаваемый согласно (5), а также графики функций $\varphi_n(x)$ (см. (6)) при $n=3$ и $n=10$.)

Наконец, можно отказаться и от условия положительности $\varphi(x)$. Очень любопытный и важный пример представляет собой

$$R(x) = \frac{1}{2\pi}\int_{-\omega}^{+\omega} \cos\xi x\,d\xi = \frac{1}{\pi}\frac{\sin\omega x}{x}. \qquad (16.4.7)$$

[3] Строго говоря, если при каких-то значениях x функция $f(x)$ разрывна и $f(x)\to\infty$ (или $f(x)\to\infty$ при $x\to\infty$ или при $x\to-\infty$), то не всякие $\varphi(x)$ годятся для получения (4). Функция $f(x)$ не должна быть разрывной или по крайней мере ее разрывы не должны приходиться на ту точку $(x=0)$, где $\delta(x)=\infty$, — иначе как раз и возникнут бессмысленные вопросы о значении функции в точке разрыва.

При заданном ω график $R(x)$ показан на рис. 2. Значение $R(0)$ равно $\frac{\omega}{\pi}$ (неопределенность, связанная с одновременным обращением числителя и знаменателя в нуль при $x=0$, раскрывается элементарно). Функция $R(x)$ принимает нулевое значение и меняет знак при $x=\pm\frac{\pi}{\omega}, \pm\frac{2\pi}{\omega}, \pm\frac{3\pi}{\omega}, \ldots$ Колебания $R(x)$ затухают с удалением от $x=0$ за счет знаменателя. Кривая не выходит за пределы линий $y=\pm\frac{1}{\pi|x|}$ (пунктир на рис. 2). Можно проверить, что $\int\limits_{-\infty}^{\infty} R(x)\,dx=1$ при любом ω.

Оказывается, что при $\omega\to\infty$ можно рассматривать $R(x)$ как дельта-функцию! Это правдоподобно потому, что с увеличением ω растет высота $\frac{\omega}{\pi}$ главного максимума, приходящегося на ось R, и уменьшается ширина этой полуволны $-\frac{\pi}{\omega}<x<\frac{\pi}{\omega}$. Но как быть с тем фактом, что при увеличении ω не уменьшается амплитуда колебаний, по-прежнему

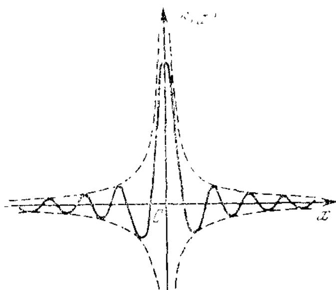


Рис. 16.4.2

R достигает $\pm\frac{1}{\pi x}$, пунктирные линии не сужаются? Рассмотрим $\int f(x)\,R(x)\,dx$. Чем больше ω, тем чаще колебания, тем точнее компенсируют одна другую положительные и отрицательные полуволны, между тем вклад первой полуволны и ближайших к ней все время одинаков. Поэтому оказывается (доказательства здесь мы не приводим), что $\lim\limits_{\omega\to\infty}\int f(x)\,R(x)\,dx=f(0)$ — а это и означает, что $\lim\limits_{\omega\to\infty} R(x)$ имеет свойства дельта-функции.

Интересна аналогичная функция

$$P(x)=\frac{1}{\pi}\left(\frac{1}{2}+\sum_{k=1}^{k=q}\cos kx\right). \qquad (16.4.8)$$

С помощью формул элементарной тригонометрии можно получить выражение

$$P(x)=\frac{1}{2\pi}\,\frac{\sin\left(q+\frac{1}{2}\right)x}{\sin\frac{x}{2}} \qquad (16.4.8a)$$

(см. упр. 1). График $P(x)$ при $q=10$ показан на рис. 3. При $q\to\infty$ функция $P(x)$ ведет

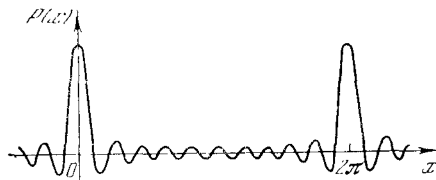


Рис. 16.4.3

себя вблизи $x=0$ совершенно так же, как $R(x)$ при $\omega\to\infty$. Отличие $P(x)$ от $R(x)$ заключается в том, что у $P(x)$ высокие максимумы повторяются периодически при $x=0$, $x=\pm2\pi$, $x=\pm4\pi$ и т. д., т. е. $P(x)$ при $q\to\infty$ представляет собой с у м м у δ-функций:

$P(x)=\delta(x)+\delta(x-2\pi)+\delta(x+2\pi)+\delta(x-4\pi)+$
$+\delta(x+4\pi)+\ldots$

Функции R и P и связь их с дельта-функцией — это не математические курьезы. Вспомним, как построены R и P: R — это интеграл косинуса, P — это сумма косинусов. Если из косинусов можно сложением построить $\delta(x)$ (напомним, что интегрирование — это разновидность сложения!), то $\delta(x-a)$ можно построить из $\cos\omega(x-a)=\cos\omega x\cos\omega a-\sin\omega x\sin\omega a$, т. е. из косинусов и синусов с постоянными коэффициентами. Но в таком случае и любую функцию $f(x)$ можно представить в виде суммы косинусов и синусов. Любую функцию можно заменить серией ступенек $f(x_i)\,\Delta x_i$, а каждая такая ступенька — это, в сущности, $\delta(x-x_i)\,f(x_i)\,\Delta x_i$. Таким образом, с помощью R и P, т. е. по существу через посредство дельта-функции, доказывается возможность разложения функций в *ряд Фурье* (если функция периодична; ср. § 10.9) и в так называемый ***интеграл Фурье*** (если функция непериодична).

Перечитайте после этого параграфа снова § 10.8 и 10.9. Обычно в учебниках математики дельта-функция не упоминается — многие педагоги предпочитают возможно дольше не знакомить учащихся с такой «физической ересью». Однако понимание того, что в доказательствах в действительности используется дельта-функция, поможет Вам глубже проникнуть в смысл этих доказательств.

Упражнение

16.4.1. Докажите совпадение правых частей формул (8) и (8а).

Глава 17

НЕКОТОРЫЕ ПРИЛОЖЕНИЯ ФУНКЦИЙ КОМПЛЕКСНОЙ ПЕРЕМЕННОЙ И ДЕЛЬТА-ФУНКЦИИ

§ 1. Комплексные числа и механические колебания

Настоящая, третья, часть книги не претендует на большее, чем лишь на то, чтобы заинтересовать читателя такими, казалось бы не похожими на «обыкновенные» функциями, как функции комплексной переменной и дельта-функции, указать на принципиальную возможность распространения аппарата высшей математики и на эти необычные объекты. И в этой последней главе книги мы лишь весьма бегло наметим некоторые из возможных путей приложения описанных в главах 14—16 функций к физическим и техническим задачам.

Как упрощаются вычисления при использовании комплексных переменных, мы покажем на примере вынужденных колебаний тяжелого тела (см. гл. 10, особенно § 10.5). Начнем с уравнения таких колебаний. Положим, что на тело массы m (на «материальную точку» массы m — размерами тела мы пренебрегаем) действует упругая (возвращающая) сила $-kx$, пропорциональная отклонению x (для наглядности можно представлять здесь пружину, стремящуюся вернуть тело в состояние равновесия), сила трения $-h\frac{dx}{dt}$, пропорциональная скорости $\frac{dx}{dt}$ (и направленная противоположно скорости), а также некоторая внешняя сила $F=F(t)$. В таком случае уравнение (9.4.2) движения тела (уравнение Ньютона; см. § 9.4) примет вид

$$m\frac{d^2x}{dt^2}=-kx-h\frac{dx}{dt}+F, \quad \text{или}$$

$$m\frac{d^2x}{dt^2}+h\frac{dx}{dt}+kx=F. \tag{17.1.1}$$

Мы в основном ограничимся случаем периодической внешней силы $F=f\cos\omega t$ (ср., впрочем, со сказанным ниже, в конце этого параграфа). Для периодической силы (с *частотой* ω и *амплитудой* f) уравнение (1) примет вид

$$m\frac{d^2x}{dt^2}+h\frac{dx}{dt}+kx=f\cos\omega t \tag{17.1.2}$$

(см. (10.5.1)); здесь m, k, h, f и ω — некоторые постоянные величины[1].

До сих пор мы считали все величины в (1) и (2) вещественными. Условимся теперь временно (ибо в окончательном ответе это предложение придется снова исключить!) допускать *комплексные* решения x нашего уравнения; все же остальные величины, являющиеся коэффициентами (2), пусть по-прежнему являются вещественными. Выражение $f\cos\omega t$ внешней силы нам теперь будет удобнее заменить на

$$fe^{i\omega t}\ (=f\cos\omega t+if\sin\omega t), \tag{17.1.3}$$

дополнив последний член уравнения (2) мнимым слагаемым $if\sin\omega t$.

Решение уравнения

$$m\frac{d^2x}{dt^2}+h\frac{dx}{dt}+kx=fe^{i\omega t} \tag{17.1.2а}$$

мы будем искать в виде

$$x=ae^{i\omega t}. \tag{17.1.4}$$

Такой вид решения подсказывается тем, что все производные экспоненциальной функции, как нам известно, пропорциональны самой этой функции; поэтому после подстановки такого выражения для x в левую и правую части уравнения (2а) в обеих его частях будет фигурировать один и тот же множитель $e^{i\omega t}$, который мы сможем просто сократить. Действительно, так как согласно (4) $\frac{dx}{dt}=ia\omega e^{i\omega t}$ и $\frac{d^2x}{dt^2}=-a\omega^2e^{i\omega t}$, то уравнение (2а) дает

$$-am\omega^2e^{i\omega t}+iha\omega e^{i\omega t}+ake^{i\omega t}=fe^{i\omega t},$$

[1] Поскольку m имеет размерность кг, а x, $\frac{dx}{dt}$ и $\frac{d^2x}{dt^2}$ — соответственно размерности м, м/с и м/с² (cos ωt, конечно, безразмерен), то ясно, что коэффициенты k, h и f уравнения (2) имеют размерности кг/с², кг/с и кг·м/с² (т. е. Н: ведь f — это сила).

или

$$a(-m\omega^2+k+ih\omega)=f,$$

т. е. $$a=\frac{f}{-m\omega^2+k+ih\omega}. \qquad (17.1.5)$$

Итак, (комплексная!) амплитуда a колебаний $x=ae^{i\omega t}$ $(=a(\cos\omega t+i\sin\omega t))$ оказалась пропорциональной приложенной силе f, что совершенно естественно для линейной системы (для решения линейного уравнения (2а)). При небольшой силе трения (т. е. небольшой величине h) максимум амплитуды (точнее, модуля амплитуды) достигается при такой частоте ω, что $m\omega^2\approx k$, т. е. при частоте ω внешней силы, близкой к собственной частоте $\omega_0=\sqrt{\frac{k}{m}}$ системы (ср. § 10.2); таким образом, здесь имеет место явление резонанса (§ 10.5).

Выражение (5) комплексной амплитуды a можно переписать так:

$$a=fre^{i\varphi}, \qquad (17.1.5а)$$

где $re^{i\varphi}=\frac{1}{-m\omega^2+k+i\omega h}$, т. е.

$$r=\frac{1}{\sqrt{(-m\omega^2+k)^2+h^2\omega^2}},$$

$$\varphi=\operatorname{arctg}\left(\frac{h\omega}{m\omega^2-k}\right). \qquad (17.1.5б)$$

Тогда

$$x=ae^{i\omega t}=fre^{i(\omega t+\varphi)}, \qquad (17.1.6)$$

где r и φ задаются формулами (5б). Таким образом, компактная комплексная запись (4) вещественного решения уравнения (2) указывает как амплитуду $|x|=fr$, так и частоту ω и фазу φ решения, точнее — сдвиг фазы колебания по отношению к фазе силы, принятой за нуль (см. (5б)).

Вернемся, наконец, к вещественным переменным. Так как наше исходное уравнение (1) — линейное, то для описываемой этим уравнением механической системы выполняется *принцип суперпозиции*: движение, вызываемое сложной силой F_1+F_2, может быть получено сложением движений, вызываемых в отдельности силами F_1 и F_2 (этим мы уже неоднократно пользовались выше). В нашем случае сила (3) распадается на вещественную компоненту $f\cos\omega t$ и мнимую компоненту $if\sin\omega t$. Но все коэффициенты уравнения (1), за исключением «свободного члена» F (члена, не содержащего неизвестной функции x), вещественны; поэтому вещественной силе F может отвечать лишь вещественное значение x, а (чисто) мнимой силе F — мнимое x. Другими словами, наша линейная система рождает вещественный отклик на вещественную внешнюю силу и мнимый отклик — на мнимую силу. Поэтому можно быть уверенным, что внешней силе $F=f\cos\omega t$ отвечает решение

$$x=fr\cos(\omega t+\varphi) \qquad (17.1.7)$$

уравнения (2), где r и φ даются формулами (5б), а силе $F=if\sin\omega t$ — решение $x=ifr\sin(\omega t+\varphi)$ (т. е. силе $F=f\sin\omega t$ уравнения (1) — решение $x=fr\sin(\omega t+\varphi)$, что, впрочем, не доставляет нам новой информации, ибо это решение получается из (7) подстановкой $t\to t+\frac{\pi}{2\omega}$).

Тот же результат можно получить и несколько по-иному. В силу формул Эйлера (14.3.4) вещественная сила $F=f\cos\omega t$ есть сумма (комплексно-сопряженных) сил $F_1={}^1/_2fe^{i\omega t}$ и $F_2={}^1/_2fe^{-i\omega t}$; при этом в силу линейности (1) (в силу принципа суперпозиции) движение, вызванное суммой F_1+F_2 сил, представляет собой сумму движений, порожденных каждой из них в отдельности. Но мы уже видели, что внешняя сила F_1 порождает движение $x={}^1/_2fre^{i(\omega t+\varphi)}$, где r и φ задаются равенствами (5б). Аналогично находится и решение, отвечающее значению F_2 силы, после чего легко найти и сумму двух решений (см. упр. 1).

Отметим в заключение некоторые свойства полученного решения, которые сохраняются даже в более общих случаях, когда рассматриваемая колебательная система состоит из многих масс, пружин («возвращающих» сил) и трущихся поверхностей (многих сил трения).

1. Мнимая часть «комплексной амплитуды» (5а) отрицательна — она равна $-\frac{(f\omega h)\,i}{r^2}$ (почему?). Это соответствует отрицательному углу φ (или углу φ, заключающемуся в пределах $\pi<\varphi<2\pi$).

Отрицательность мнимой части a связана с работой, которую производит сила F в среднем за время многих колебаний или точно за один цикл. Если (внешняя) сила задается равенством $F=f\cos\omega t$, а решение имеет вид (7), то работа силы равна

$$\int F dx = \int F \frac{dx}{dt} dt =$$
$$= \int f \cos(\omega t) [-\omega f r \sin(\omega t + \varphi)] dt =$$
$$= -f^2 r\omega \int \sin(\omega t + \varphi) \cos(\omega t) dt = \qquad (17.1.8)$$
$$= -\frac{1}{2} f^2 r\omega \int [\sin(2\omega t + \varphi) + \sin\varphi] dt =$$
$$= \frac{1}{2} f^2 r\omega \left[- \int \sin\varphi dt - \int \sin(2\omega t + \varphi) dt \right].$$

Ясно, что второй интеграл в правой части (8), распространенный на время τ, отвечающее одному циклу колебаний $\left(\tau = T = \frac{2\pi}{\omega}\right)$, обратится в нуль, так что работа будет равна $-{}^1/_2 f^2 r\omega (\sin\varphi)\tau$. Но эта работа должна быть положительной, поскольку за рассмотренное время система вернется в прежнее положение, отвечающее тем же значениям и кинетической и потенциальной энергии; поэтому на смещение системы работа силы F затрачена не будет и вся она уйдет на нейтрализацию работы трения, которая всегда является отрицательной: ведь сила трения по направлению противоположна скорости $\frac{dx}{dt}$, а значит, и смещению dx.

Заметим, что за время, не равное полному циклу колебаний, например за короткое время, много меньшее, чем T, работа силы F может быть и положительной, и отрицательной, ибо на этот краткий период времени может приходиться уменьшение энергии тела (суммы его потенциальной и кинетической энергий). Определенное заключение о знаке работы можно сделать, лишь если речь идет об очень длительном промежутке времени функционирования системы (или о времени одного цикла — большой промежуток времени складывается из многих однотипных периодов (циклов)).

2. Величину a (см. (5)—(5б)) можно рассматривать как функцию комплексной частоты ω. При этом a обращается в бесконечность при значении $\omega = \omega_p$ (ω_p — резонансная частота), где $\omega_p = \lambda + i\mu$ обязательно имеет положительную мнимую часть (т. е. обязательно $\mu > 0$). Бесконечность a, т. е., точнее, бесконечность отношения $\frac{a}{f}$, означает, что $\frac{f}{a} = 0$, т. е. что на частоте ω_p возможны свободные колебания, не требующие никакой внешней силы F. Но в таком случае решение $x(t) = ae^{i\omega_p t}$ удовлетворяет уравнению

$$m\frac{d^2x}{dt^2} + h\frac{dx}{dt} + kx = 0, \qquad (17.1.2б)$$

получаемому из (1) подстановкой $F = 0$, откуда вытекает, что

$$-m\omega_p^2 + ih\omega_p + k = 0 \qquad (17.1.9)$$

(ср. со знаменателем в формуле (5) для a). Но ясно, что подобные свободные колебания (в условиях наличия трения!) могут быть только затухающими. Поэтому в решении $e^{i\omega_p t} = e^{i(\lambda + i\mu)t} = e^{(-\mu + i\lambda)t}$ величина $-\mu$ должна быть отрицательной, т. е. должно быть $\mu > 0$.

В нашем случае квадратное уравнение (9) дает

$$\omega_p = \frac{ih}{2m} \pm \sqrt{\left(\frac{ih}{2m}\right)^2 + \frac{k}{m}} =$$
$$= \pm \sqrt{\frac{k}{m} - \left(\frac{h}{2m}\right)^2} + \frac{ih}{2m}.$$

Если здесь $\frac{k}{m} - \frac{h^2}{4m^2} = \varkappa^2 > 0$, то мы приходим к двум решениям $\omega_{p1} = \varkappa + \frac{ih}{2m}$ и $\omega_{p2} = -\varkappa + \frac{ih}{2m}$ с одной и той же (положительной!) мнимой частью $\frac{h}{2m} i$. (Эти решения являются «почти комплексно-сопряженными», точнее, $\omega_{p1} = -\omega_{p2}^*$ и $\omega_{p2} = -\omega_{p1}^*$.) Если же $\frac{k}{m} - \frac{h^2}{4m^2} < 0$, т. е. $\frac{h^2}{4m^2} - \frac{k}{m} = \varkappa_1^2 > 0$, то наши решения имеют вид $\omega_{p1} = \left(\frac{h}{2m} + \varkappa_1\right) i$ и $\omega_{p2} = \left(\frac{h}{2m} - \varkappa_1\right) i$ — это два чисто мнимых числа, имеющих положительные коэффициенты при i, так как $\varkappa_1 < \frac{h}{2m}$.

Те же методы, базирующиеся на использовании комплексной переменной x, применимы и в более общем случае произвольной, а не только синусоидальной силы F. В случае произвольной периодической силы F нам достаточно разложить эту силу в ряд Фурье (см. § 10.9) и затем воспользоваться принципом суперпозиции, рассматривая в отдельности действие, которое производит каждая из составляющих синусоидальных сил (3); при этом удобна комплексная форма ряда Фурье, т. е. представление произвольной периодической функции $F(t)$ с периодом T в виде линейной комбинации функций $e^{ni\omega t}$, где $\omega = 2\pi/T$ (основная частота; ср. (10.2.5)), а $n = \ldots, -2, -1, 0, 1, 2, \ldots$ — целое, т. е. в виде суммы функций $e^{in\omega t}$, взятых с какими-то коэф-

фициентами f_n. В случае непериодической силы F разложение ее в ряд Фурье приходится заменять разложением в так называемый интеграл Фурье, распространенный по всевозможным значениям частоты ω (а не только по частотам $\omega_n = n\omega$, кратным основной частоте ω, как в случае ряда Фурье); при этом и для интеграла Фурье наиболее удобной будет комплексная его форма.

Не останавливаясь здесь на деталях соответствующих построений (см., впрочем, упр. 2), укажем, что возможен и иной подход к задаче о колебании, возбужденном произвольной силой $F=F(t)$, а именно с помощью разбиения силы F на отдельные дельта-образные слагаемые, отвечающие узкой полоске графика функции F, — замены этой функции на сумму функций $F_\tau(t)$, где

$$F_\tau(t) = \begin{cases} F(t) & \text{при } \tau \leqslant t \leqslant \tau + d\tau, \\ 0 & \text{при всех других } t, \end{cases} \quad (17.1.10)$$

— и последующему использованию принципа суперпозиции. (Решение, отвечающее дельта-образной силе $F_\tau(t)$, называется *функцией Грина* соответствующего дифференциального уравнения; см. по этому поводу § 4.) При этом мы приходим к решению уравнения (1), имеющему в более простом случае отсутствия сил трения (в случае $h=0$) вид

$$x(t) = A \int_{-\infty}^{t} e^{i\omega(t-\tau)} F(\tau)\, d\tau, \quad (17.1.11)$$

где A и ω определяются коэффициентами m и k уравнения (1), или, в «вещественной» форме,

$$x(t) = A \int_{-\infty}^{t} \sin[\omega(t-\tau)] F(\tau)\, d\tau, \quad (17.1.11а)$$

где $\omega = \sqrt{\frac{k}{m}}$ (ср. § 10.2) и $A = \sqrt{km}$.

По поводу вывода формул (11) или (11а) см. § 4; в частности ср. (11а) и (4.5а). Нетрудно также, не вдаваясь в соображения, с помощью которых формула (11а) была получена, непосредственно проверить, что она правильно указывает решение уравнения (1) для случая $h=0$. В самом деле, функция, стоящая в правой части равенства (11а), двояким образом зависит от t: во-первых, t есть верхний предел интеграла; во-вторых, от t (в комбинации $\sin[\omega(t-\tau)]$) зависит подынтегральная функция. Поэтому производная $\frac{dx}{dt}$ должна состоять из двух членов: первый получается при дифференцировании интеграла по верхнему его пределу, а второй — при дифференцировании по t подынтегральной функции[2]. Но производная интеграла по верхнему его пределу в силу теоремы Ньютона—Лейбница равна подынтегральной функции от верхнего предела (см. § 3.3); таким образом,

$$\frac{dx}{dt} = A \sin 0 \cdot F(t) + A \int_{-\infty}^{t} \omega \cos[\omega(t-\tau)] \times F(\tau)\, d\tau = A\omega \int_{-\infty}^{t} \cos[\omega(t-\tau)] F(\tau)\, d\tau \quad (17.1.12)$$

---

[2] Это следует из того, что если $y = F(x) = \int_a^x f(x, \tau)\, d\tau$, то

$$\Delta y = F(x+\Delta x) - F(x) = \int_a^{x+\Delta x} f(x+\Delta x, \tau)\, d\tau - \int_a^x f(x, \tau)\, d\tau = \left[\int_a^{x+\Delta x} f(x+\Delta x, \tau)\, d\tau - \int_a^x f(x+\Delta x, \tau)\, d\tau\right] + \left[\int_a^x f(x+\Delta x, \tau)\, d\tau - \int_a^x f(x, \tau)\, d\tau\right].$$

Таким образом, отношение $\frac{\Delta y}{\Delta x}$ распадается на сумму двух дробей, предел первой из которых представляет собой производную от интеграла $\int_a^x f(x+\Delta x, \tau)\, d\tau$ по верхнему его пределу, равный функции $f(x+\Delta x, \tau)$ при $\tau = x$ (которая при $\Delta x \to 0$ переходит в $f(x, x)$); вторая же дробь равна

$$\frac{1}{\Delta x} \int_a^x [f(x+\Delta x, \tau) - f(x, \tau)]\, d\tau = \int_a^x \frac{f(x+\Delta x, \tau) - f(x, \tau)}{\Delta x}\, d\tau$$

— и при естественных условиях ее предел равен $\int_a^x \frac{\partial f(x, \tau)}{\partial x}\, d\tau$.

и, следовательно,

$$\frac{d^2x}{dt^2} = A\omega \cos 0 \cdot F(t) +$$

$$+ A\omega \int_{-\infty}^{t} -\omega \sin[\omega(t-\tau)] F(\tau)\, d\tau =$$

$$= A\omega F(t) - A\omega^2 \int_{\infty}^{t} \sin[\omega(t-\tau)] F(\tau)\, d\tau =$$

$$= A\omega F(t) - \omega^2 x(t). \qquad \textbf{(17.1.13)}$$

Поэтому если, $\omega^2 = \frac{k}{m}$, т. е. $\omega = \sqrt{\frac{k}{m}}$, и $A\omega = \frac{1}{m}$, т. е. $A = \frac{1}{\omega m} = \frac{1}{\sqrt{km}}$, то

$$\frac{d^2x}{dt^2} = -\frac{k}{m}x + \frac{1}{m}F(t)$$

— а это и есть уравнение (1) с $h = 0$.

Наконец, укажем, что эти же методы «комплексного анализа» применимы и в более сложных задачах, скажем относящихся к процессу установления колебаний, где $x=0$ при $t < t_0$ и $x \neq 0$ лишь при $t > t_0$, или к случаю нескольких колеблющихся тел, когда в системе действует ряд «возвращающих сил» («пружин») и имеется много трущихся поверхностей (много сил трения, пропорциональных производным $\frac{dx}{dt}$). В теории электрических цепей (см. гл. 13) мы также постоянно встречаемся с колебательными процессами — в родственных (1) уравнениях роль коэффициентов m, h и k здесь играют индуктивность L, сопротивление R и величина $\frac{1}{C}$, обратная емкости C, а роль «внешней силы» F играет подаваемое в сеть напряжение (ЭДС). И здесь методы, аналогичные намеченным выше (рассмотрение комплексных решений вещественных дифференциальных уравнений), используются весьма широко — настолько широко, что с понятием «мнимого тока» знаком буквально каждый, кто когда-либо изучал электротехнику.

Мы здесь не будем останавливаться подробнее на всех этих вопросах.

Упражнения

17.1.1. Получите решение (7) уравнения (2) как сумму двух решений уравнения (1), отвечающих значениям $F_1 = {}^1/_2 fe^{i\omega t}$ и $F_2 = {}^1/_2 fe^{-i\omega t}$ силы F.

17.1.2. Докажите, что заданная на интервале от $-\pi$ до π (вещественная!) функция $f(x)$ может быть представлена в виде суммы (комплексная форма ряда Фурье):

$$f(x) = \sum_{-\infty}^{\infty} c_k e^{ikx} \; (= \ldots + c_{-2}e^{-2ix} +$$
$$+ c_{-1}e^{-ix} + c_0 + c_1 e^{ix} + c_2 e^{2ix} + \ldots),$$

где $c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\, e^{-ikt} dt$ (здесь $k = \ldots -2, -1, 0, 1, 2, \ldots$) и $c_{-k} = c_k^*$; если же значения $f(x)$ — комплексные числа (но x по-прежнему вещественно), то имеют место все те же формулы, где только теперь уже не обязательно $c_{-k} = c_k^*$.

§ 2. Интегралы в комплексной области

Рассмотрим определенный интеграл

$$w(z) = \int_a^z f(z)\, dz, \qquad \textbf{(17.2.1)}$$

где нижним пределом служит фиксированное комплексное число a, а верхний предел z — переменный. Интеграл здесь понимается так же, как и в области вещественной переменной (см. гл. 3), т. е. правая часть (1) приближенно равна сумме

$$f(z_0)\Delta z_0 + f(z_1)\Delta z_1 + \ldots$$
$$\ldots + f(z_{n-1})\Delta z_{n-1} = \sum_{i=0}^{n-1} f(z_i)\Delta z_i,$$

где $z_0 = a$, z_1, z_2, ..., z_{n-1}, $z_n = z$ — точки, разбивающие (соединяющий a и z) путь C интегрирования на достаточно малые части, а $\Delta z_i = z_{i+1} - z_i$; $i = 0, 1, \ldots, n-1$.

По какому пути C следует проводить интегрирование? На рис. 1 показаны для примера три таких пути. Однако можно доказать, что *если в области, ограниченной двумя путями C и C_1, функция не имеет «особенностей»*, т. е. она всюду имеет смысл и нигде не обращается в бесконечность, *то интеграл* (1), *взятый по этим путям, имеет одно и то же значение*. Разумеется, само упоминание о производной $f'(z)$ функции $f(z)$ комплексной переменной подразумевает, что эту функцию мы здесь, как и всегда, считаем, аналитической.

Сформулированное свойство интегралов (1) допускает еще одну формулировку, эквивалентную первоначальной.

Обозначим символом $\oint f(z)\,dz$ интеграл, распространенный по *замкнутому* контуру, скажем по контуру, получающемуся, если мы пройдем от точки a к точке z по пути C, а затем вернемся из z в точку a по пути $-C_1$ (путь C_1 соединяет a с z, мы же идем по той же дороге в обратном направлении; см. рис. 1). Так как (ср. § 3.2) $\int\limits_{-C_1} f(z)\,dz = -\int\limits_{C_1} f(z)\,dz$, то

$$\oint f(z)\,dz = \int\limits_{C} f(z)\,dz + \int\limits_{-C_1} f(z)\,dz = \int\limits_{C} f(z)\,dz - \int\limits_{C_1} f(z)\,dz,$$

и если $\int\limits_{C_1} f(z)\,dz = \int\limits_{C} f(z)\,dz$, то $\oint f(z)\,dz = 0$: *интеграл* $\oint f(z)\,dz$ *от аналитической функции по замкнутому контуру* всегда *равен нулю, если внутри этого контура функция не имеет особенностей.* Обратно, разумеется, из равенства нулю интеграла по любому замкнутому контуру следует независимость его от пути интегрирования.

Чтобы доказать сформулированное утверждение, рассмотрим вначале маленький квадратик с вершинами *1* (x_0, y_0); *2* $(x_0+\Delta, y_0)$; *3* $(x_0+\Delta, y_0+\Delta)$; *4* $(x_0, y_0+\Delta)$ (рис. 2). Символом $\oint f(x)\,dz$ мы обозначим интеграл по замкнутому контуру, ведущему из точки *1* последовательно через точки *2*, *3* и *4* снова в ту же точку *1*; вычислим этот интеграл.

Пусть $f(z) = f(x+iy) = u(x, y) + iv(x, y)$; величину Δ мы положим настолько малой, что можно считать

$$\oint f(z)\,dz = \int\limits_1^2 f(z)\,dz + \int\limits_2^3 f(z)\,dz + \int\limits_3^4 f(z)\,dz + \int\limits_4^1 f(z)\,dz \approx f(z_1)\,\Delta z_1 + f(z_2)\,\Delta z_2 + f(z_3)\,\Delta z_3 + f(z_4)\,\Delta z_4,$$

где под z_j понимается точка j, а $\Delta z_j = z_{j+1} - z_j$; здесь $j = 1, 2, 3, 4$ и $z_{4+1} = z_1$ — снова точка *1*. Так как $\Delta z_1 = z_2 - z_1 = \Delta = -\Delta z_3 = z_3 - z_4$ и $\Delta z_2 = z_3 - z_2 = i\Delta = -\Delta z_4 = z_4 - z_1$, то

$$\oint f(z)\,dz \approx \{[u(x_0, y_0) + iv(x_0, y_0)] - [u(x_0+\Delta, y_0+\Delta) + iv(x_0+\Delta, y_0+\Delta)]\}\,\Delta + \{[u(x_0+\Delta, y_0) + iv(x_0+\Delta, y_0)] - [u(x_0, y_0+\Delta) + iv(x_0, y_0+\Delta)]\}\,i\Delta =$$

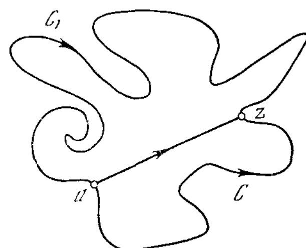


Рис. 17.2.1

$$= -\{[u(x_0+\Delta, y_0+\Delta) - u(x_0, y_0)] + [v(x_0+\Delta, y_0) - v(x_0, y_0+\Delta)]\}\,\Delta - \{[v(x_0+\Delta, y_0+\Delta) - v(x_0, y_0)] + [u(x_0, y_0+\Delta) - u(x_0+\Delta, y_0)]\}\,i\Delta = -A\cdot\Delta - B\cdot i\Delta,$$

где через A и B обозначены выражения, заключенные в фигурные скобки. Но

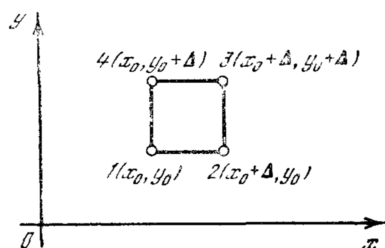


Рис. 17.2.2

$$u(x_0+\Delta, y_0+\Delta) - u(x_0, y_0) = [u(x_0+\Delta, y_0+\Delta) - u(x_0, y_0+\Delta)] + [u(x_0, y_0+\Delta) + u(x_0, y_0)] \approx \left[\frac{\partial u(x_0, y_0+\Delta)}{\partial x} + \frac{\partial u(x_0, y_0)}{\partial y}\right]\Delta$$

и

$$v(x_0+\Delta, y_0) - v(x_0, y_0+\Delta) = [v(x_0+\Delta, y_0) - v(x_0+\Delta, y_0+\Delta)] + [v(x_0+\Delta, y_0+\Delta) - v(x_0, y_0+\Delta)] \approx \left[-\frac{\partial v(x_0+\Delta, y_0)}{\partial y} + \frac{\partial v(x_0, y_0+\Delta)}{\partial x}\right]\Delta,$$

где мы по-прежнему пользуемся приближенной формулой: $f(x+\Delta) - f(x) \approx f'(x)\,\Delta$. Таким образом,

$$A \approx \left\{\left[\frac{\partial u(x_0,\ y_0+\Delta)}{\partial x} - \frac{\partial v(x_0+\Delta,\ y_0)}{\partial y}\right] + \right.$$
$$\left. + \left[\frac{\partial u(x_0,\ y_0)}{\partial y} + \frac{\partial v(x_0,\ y_0+\Delta)}{\partial x}\right]\right\}\Delta \approx$$
$$\approx \left\{\left[\frac{\partial u}{\partial x} - \frac{\partial v}{\partial y}\right] + \left[\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right]\right\}\Delta = 0$$

в силу соотношений Коши—Римана. (В последних выражениях аргументы всех функций $\frac{\partial u}{\partial x}$, $\frac{\partial v}{\partial y}$, $\frac{du}{dy}$, $\frac{dv}{dx}$ можно приравнять x_0, y_0, т. е. считать, что частные производные вычисляются в вершине изображенного на рис. 2 квадратика — при этом мы совершаем ошибку порядка Δ.) Аналогично устанавливается и равенство нулю выражения B.

Разумеется, на самом деле мы установили лишь, что наш интеграл п о ч т и равен нулю, что A и B отличаются от нуля на величину порядка малости выше Δ, а сам интеграл — на величину порядка малости выше площади Δ^2 квадратика (порядка Δ^3 или Δ^4, или еще меньшую). Отсюда, однако, уже следует, что наш интеграл равен нулю.

В самом деле, разобьем исходный квадрат со стороной Δ на n^2 меньших квадратов со стороной $\frac{\Delta}{n}$. Интеграл, распространенный по контуру исходного квадрата, равен, как легко понять, сумме n^2 аналогичных интегралов, распространенных по контурам всех меньших квадратов (где все контуры мы обходим в одном и том же направлении, скажем против часовой стрелки): ведь при суммировании всех интегралов каждый внутренний отрезок окажется пройденным 2 раза в двух противоположных направлениях и сумма этих слагаемых сократится. Но если большой интеграл имеет величину порядка Δ^3, то маленькие квадраты имеют величину порядка $(\Delta/n)^3$ каждый, а сумма таких интегралов — величину порядка $n^2\left(\frac{\Delta}{n}\right)^3 = \frac{\Delta^3}{n}$. Так как эта оценка справедлива при л ю б о м n, то наш интеграл неизбежно должен равняться нулю.

Таким же способом, т. е. разбивая область внутри любого замкнутого контура на малые квадратики, можно доказать равенство нулю интеграла по контуру этой области и тем самым — независимость интеграла (1) от пути интегрирования.

П р и м е р. Рассмотрим интеграл

$$I = \int_0^{1+i} z^2 dz. \qquad (17.2.2)$$

Пусть путь A ведет из точки $z_0=0$ до точки $z=1+i$ сначала по вещественной оси (от $z_0=0$ до $z_1=1$), а затем по прямой, параллельной мнимой оси (от $z_1=1$ до $z=1+i$). При этом имеем

$$I = \int_A z^2 dz = \int_0^1 (x^2 - y^2 + 2ixy)|_{y=0} dx +$$
$$+ \int_0^1 (x^2 - y^2 + 2ixy)|_{x=1} i dy = -\frac{2}{3} + \frac{2}{3} i.$$

Тот же численный результат получится, если мы выберем другой путь B — сначала по мнимой оси от точки $z_0=0$ до точки $z_2=i$, а затем по прямой, параллельной вещественной оси (от точки $z_2=i$ до конечной точки $z=1+i$; проверьте!). Наконец, можно интегрировать и по диагонали C натянутого на точки z_0, z_1, z, z_2 прямоугольника (здесь — квадрата): положим $z=(1+i)t$, где $0 \leqslant t \leqslant 1$; тогда $dz=(1+i)dt$ и

$$I = \int_0^{1+i} z^2 dz = \int_0^1 [(1+i)t]^2 (1+i) dt =$$
$$= (1+i)^3 \int_0^1 t^2 dt = \frac{1}{3}(1+i)^3 = -\frac{2}{3} + \frac{2}{3} i.$$

Значение интеграла (2) можно получить и без всех этих немудрящих арифметических упражнений. Для этого, как и в случае интегрирования вещественных функций, достаточно найти п е р в о о б р а з н у ю функцию $F(z)$ функции z^2. Пусть $F(z)$ — такая (разумеется, аналитическая!) функция комплексной переменной, что

$$\frac{dF}{dz} = f(z). \qquad (17.2.3)$$

Тогда в силу теоремы Ньютона—Лейбница

$$\int_a^z f(z)dz = F(z) + C,$$
$$I(a,\ b) = \int_a^b f(z)dz = F(b) - F(a), \qquad (17.2.4)$$

где a, b, C, $F(a)$, $F(b)$ — комплексные числа. Равенства (4) доказываются совершенно так же, как для функций вещественной переменной — находим изменение интеграла $I(a,\ z) = \int_a^z f(z)dz$

при малом изменение верхнего предела интегрирования:

$$I(a, z+\Delta z) - I(a, z) = \int_{z}^{z+\Delta z} f(z)dz = \\ = f(z)\Delta z + \ldots \qquad (17.2.5)$$

(здесь опущены слагаемые, имеющие (по модулю) величину порядка $|\Delta z|^2$ *или* еще меньшую), откуда в пределе при малом (по абсолютной величине) Δz (причем независимо от «направления» приращения Δz, т. е. от отношения вещественной его части Δu к коэффициенту Δv при мнимой части) получаем

$$\frac{\Delta F(z)}{\Delta z} \approx f(z). \qquad (17.2.6)$$

Из (6) вытекает, что значение интеграла $I(a, z)$ может отличаться от функции $F(z)$ только на константу. Значение этой константы находим из условия $I(a, z)=0$ при $z=a$, откуда $I(a, z)=F(z)-F(a)$ в соответствии с (4). Но так как (ср. с. 427) и в комплексной области $(z^3)'=3z^2$, то $\int z^2dz=\frac{z^3}{3}+C$, — откуда и вытекает полученное выше значение интеграла (2).

Мы здесь изложили все эти соображения столь бегло потому, что в этом пункте отсутствуют какие бы то ни было различия между (гладкими) функциями вещественной переменной и (аналитическими) функциями комплексной переменной — все результаты гл. 3 дословно переносятся на случай функций комплексной переменной.

Перейдем теперь к специфическим свойствам интегралов от функций комплексной переменной, не имеющим аналогов в вещественной области. Обратимся к интегралу

$$\int \frac{1}{z}dz \quad (=\ln z). \qquad (17.2.7)$$

В качестве пути интегрирования в (7) выберем окружность фиксированного радиуса ρ с центром в точке $z=0$, т. е. положим $z=\rho e^{i\varphi}$, где $\rho=\text{const}$ и $0\leqslant\varphi\leqslant 2\pi$; тогда $dz=i\rho e^{i\varphi}d\varphi$ и $\frac{dz}{z}=id\varphi$ (равенство $d(e^{i\varphi})=ie^{i\varphi}d\varphi$ вытекает из того, что по определению величины $e^{i\varphi}$ при малом $\Delta\varphi$ имеем $e^{i\Delta\varphi}\approx 1+i\Delta\varphi$).

Поэтому интеграл (7) по нашему замкнутому контуру оказывается равным

$$\oint \frac{1}{z}dz = \int_0^{2\pi} id\varphi = 2\pi i, \qquad (17.2.8)$$

т. е. он отнюдь не равен нулю. При этом значение (8) интеграла не зависит от радиуса ρ круга; более того, оно не зависит и от формы охватывающего точку 0 контура — тот же результат $2\pi i$ мы получим и при интегрировании по любому другому замкнутому контуру, содержащему 0 внутри себя, скажем по границе квадрата с вершинами $\pm 1 \pm i$ (см. упр. 1).

Почему же основной результат о равенстве нулю интеграла $\oint f(z)dz$ здесь оказывается неверным? Дело в том, что в данном случае подынтегральная функция $f(z)=\frac{1}{z}$ внутри контура интегрирования имеет особенность: в точке $z=0$ она обращается в бесконечность. При этом как бы мы ни дробили область внутри контура на отдельные малые подобласти, значение $f(z)=\infty$ неминуемо попадает внутрь одной из областей — и оно даст свой вклад: шила в мешке и бесконечность внутри контура интегрирования не утаишь!

Рассмотрим подробнее наш интеграл (7). Этот интеграл, взятый по контуру, окружающему точку 0, равен $2\pi i$. Но и при интегрировании функции $\frac{1}{z}$ по любому другому пути мы не можем игнорировать наличие у подынтегральной функции особой точки.

Пусть

$$I = \int_1^2 \frac{dz}{z} = \ln z\,|_1^2. \qquad (17.2.9)$$

Интеграл (9), если взять кратчайший путь интегрирования A, ведущий из точки $z=1$ в точку $z=2$ вдоль вещественной оси (рис. 3), равен $\ln 2\approx 0{,}69$. Но из точки $z=1$ в точку $z=2$ можно попасть и другим путем. Обойдем, например, исходя из точки $z=1$, окружность C радиуса 1 с центром в точке 0 и лишь затем пройдем по отрезку A вещественной оси от точки 1 к точке 2.

При этом мы получим другое значение интеграла (9):

$$\int\limits_1^2 \frac{dz}{z} = \int\limits_C \frac{dz}{z} + \int\limits_A \frac{dz}{z} = 2\pi i + \ln 2 \approx 0{,}69 + 6{,}28i.$$

Можно сделать не один, а n оборотов вдоль окружности C и лишь затем пройти отрезок A вещественной оси;

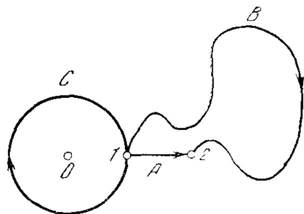


Рис. 17.2.3

при этом значение интеграла (9) окажется равным

$$\int\limits_1^2 \frac{dz}{z} = \int\limits_{nC} \frac{dz}{z} + \int\limits_A \frac{dz}{z} = \int\limits_A \frac{dz}{z} + n\int\limits_C \frac{dz}{z} = \ln 2 + 2n\pi i,$$

где под nC мы понимаем контур, получаемый при n-кратном обходе окружности C в одном и том же направлении (против часовой стрелки). С другой стороны, ту же окружность C можно проходить и в обратном направлении — по часовой стрелке; если пройти этот путь m раз, то мы получим значение $\ln 2 - 2m\pi i$ интеграла (9). При этом к значению $\ln 2 \approx 0{,}69$ интеграла (9) приводят множество путей интегрирования, а именно все пути, не обходящие точку $z=0$, вроде изображенного на рис. 3 пути B; результат $I = \int\limits_1^2 \frac{dz}{z} = \ln 2 - 4\pi i$ также дают многие (другие) пути и т. д.

Итак, полный ответ:

$$I = \int\limits_1^2 \frac{dz}{z} = \ln 2 + i2\pi k,$$

где $k = 0, \pm 1, \pm 2, \ldots$

Можно сказать, что это и есть наиболее общее определение натурального логарифма 2, тогда как число 0,69 есть только одно из многих значений величины $\ln 2$ — простейшее ее значение. В самом деле

$$e^I = e^{\ln 2 + i2\pi k} = 2e^{i2\pi k} = 2.$$

Таким образом, мы выполнили обещание, данное раньше, в § 14.3, по-новому объяснив причину неоднозначности логарифма $w = \ln z$. Эта неоднозначность связана, с одной стороны, с периодичностью функции $e^{i\varphi}$, т. е. с тем, что эта функция при $\varphi = 2\pi k$ обращается в 1. С другой стороны, логарифм можно определить как интеграл $\int \frac{dz}{z}$, — и при таком подходе неоднозначность логарифма связана с поведением функции $\frac{1}{z}$ при $z=0$ и с возможностью обхода при интегрировании этой точки $z=0$, в которой функция $f(z) = \frac{1}{z}$ обращается в бесконечность (такая точка называется *полюсом* функции комплексной переменной).

Заметим, что именно функция $\frac{1}{z}$ дает при интегрировании по обходящему точку $z=0$ контуру отличный от нуля результат $2\pi i$ (а функция $\frac{c}{z}$, где c есть фиксированное, вообще говоря, комплексное число, дает при подобном интегрировании результат $c \cdot 2\pi i$). Ясно, что интеграл по любому замкнутому контуру от функции $f(z) = a$, где a — постоянное число, или от функции $f(z) = az^n$, где степень $n > 0$ (например, от функций $2z^2$ или $-iz^4$), равен нулю — ведь эти функции вовсе не имеют никаких особенностей (полюсов). Но неожиданным образом также и отрицательные степени z, большие единицы по абсолютной величине, такие, как $\frac{1}{z^2}$ или $\frac{1}{z^3}$, тоже ничего не дают при интегрировании по замкнутому контуру, охватывающему полюс функции — точку $z=0$, в которой функция обращается в бесконечность: для таких функций $f(z)$ мы снова имеем $\oint f(z)\,dz = 0$. В самом деле, подставим в интеграл $\int \frac{dz}{z^2}$ значения $z = \rho e^{i\varphi}$, $dz = i\rho e^{i\varphi} d\varphi$, где $\rho = \text{const}$ и $0 \leqslant \varphi \leqslant 2\pi$; мы получим

$$\oint \frac{dz}{z^2} = i\int\limits_0^{2\pi} \frac{1}{\rho} e^{-i\varphi}\,d\varphi = \frac{i}{\rho(-i)} e^{-i\varphi}\Big|_0^{2\pi} = 0.$$

Так же просто доказывается равенство $\oint \frac{dz}{z^n} = 0$, где целое число $n > 1$.

Итак, значение интеграла по замкнутому контуру зависит не просто от того,

обращается ли функция в бесконечность внутри этого контура, а именно от наличия или отсутствия в выражении для функции $f(z)$ (в разложении ее в ряд, распространенный как по положительным, так и по отрицательным степеням $z-a$, где a — особая точка функции, в которой $f(z)=\infty$) члена, пропорционального $\frac{1}{z-a}$, и от коэффициента при этом члене [1]. В соответствии с особой важностью коэффициента при $(z-a)^{-1}$ в разложении $f(z)$ по степеням $(z-a)$ этому коэффициенту дали специальное название — он называется *вычетом* функции $f(z)$ в точке $z=a$. Ясно, что в точке аналитичности функции, в которой она принимает конечное значение и в окрестности которой разлагается в ряд Тейлора, вычет функции равен нулю; но также и функция $f(z)=\frac{1}{z^2}$, обращающаяся в нуле в бесконечность, имеет в этой точке вычет 0. В противоположность этому функция

$$\varphi(z)=\frac{1}{z^2+1}=\frac{1}{(z+i)(z-i)}=$$

$$=\frac{1}{2i}\frac{1}{z-i}-\frac{1}{2i}\frac{1}{z+i} \qquad (17.2.10)$$

имеет две особые точки: $z=i$ и $z=-i$; при этом вычет функции в точке $z=i$ равен $\frac{1}{2i}=-\frac{i}{2}$, а вычет в точке $z=-i$ равен $-\frac{1}{2i}=\frac{i}{2}$. В самом деле, ведь, скажем, разложение $-\frac{1}{2i}\frac{1}{z+i}$ по степеням $z-i$ отрицательных степеней не имеет, так что «общий» коэффициент при $(z-i)^{-1}$ здесь равен $\frac{1}{2i}$.

Интегралы от функций комплексной переменной имеют замечательные приложения к вычислению вещественных интегралов. Вот достаточно типичный (и достаточно важный) пример:

$$I=\int_{-\infty}^{\infty}\frac{\cos\omega x}{1+x^2}dx. \qquad (17.2.11)$$

I есть интеграл от произведения быстро меняющейся функции $\cos\omega x$ (где мы будем полагать, что $\omega \gg 1$) на плавную функцию $y_1=\frac{1}{1+x^2}$. График подынтегральной функции приведен на рис. 4, *а*, где пунктиром изображены графики функций $y_1=\frac{1}{1+x^2}$ и $y_2=-\frac{1}{1+x^2}$.

Согласно формулам Эйлера (14.3.4),

$$\cos\omega x=\frac{1}{2}e^{i\omega x}+\frac{1}{2}e^{-i\omega x}. \qquad (17.2.12)$$

Подставив (12) в правую часть (11), мы разобьем I на два интеграла I_+ и I_-, где

$$I_+=\frac{1}{2}\int_{-\infty}^{\infty}\frac{e^{i\omega z}}{1+z^2}dz. \qquad (17.2.13)$$

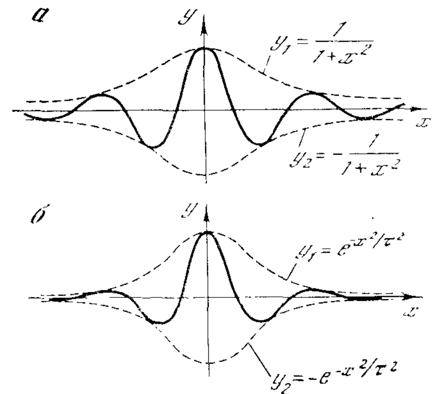


Рис. 17.2.4

Второй интеграл I_- записывается аналогично, только $e^{i\omega z}$ заменено на $e^{-i\omega z}$. При этом (немую!) переменную интегрирования в (11) мы обозначим теперь не через x, а через z, символизируя этим, что она может принимать и комплексные значения (это — центральный пункт всего рассуждения!).

Пусть z пробегает расположенную в верхней полуплоскости ($z=x+iy$, где $y>0$) полуокружность γ: $|z|=R$, где R очень велико. Так как

$$|f(z)|=\frac{1}{2}\left|\frac{e^{i\omega z}}{1+z^2}\right|=\frac{1}{2}\frac{|e^{i\omega z}|}{|1+z^2|}$$

и вдоль γ

$$|e^{i\omega z}|=|e^{i\omega(x+iy)}|=|e^{-\omega y}||e^{i\omega x}|=$$
$$=|e^{-\omega y}|\cdot 1<1$$

(ведь в верхней полуплоскости $y>0$ и $e^{-\omega y}<1$), а

$$\left|\frac{1}{1+z^2}\right|\approx\left|\frac{1}{z^2}\right|=\frac{1}{R^2},$$

то вдоль γ модуль функции $f(z)=\frac{e^{i\omega z}}{2(1+z^2)}$ будет всюду весьма мал; поэтому мал

[1] Мы не рассматриваем более сложные случаи, когда сама подынтегральная функция является неоднозначной (наподобие, скажем, функции $f(z)=\ln z$).

будет и распространенный по γ интеграл[2] $\int\limits_{\gamma} f(z)\,dz$. А отсюда следует, что при большом R интеграл

$$\frac{1}{2}\oint\limits_{\Gamma}\frac{e^{i\omega z}}{1+z^2}\,dz, \qquad (17.2.14)$$

распространенный по замкнутому контуру Г, образованному отрезком $-R \leqslant x \leqslant R$ оси x, где R очень велико, и расположенной в верхней полуплоскости полуокружностью γ (рис. 5), будет очень близок к интегралу I_+, распространенному по вещественной оси. Но интеграл от аналитической функции $f(z)$ по замкнутому контуру зависит, как мы знаем, лишь от особых точек функции $f(z)$, расположенных внутри этого контура. А есть ли внутри

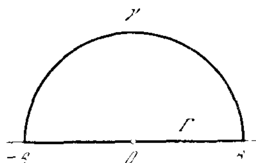


Рис. 17.2.5

нашего контура у функции $f(z) = \frac{e^{i\omega z}}{1+z^2} = e^{i\omega z}\varphi(z)$ особые точки?

Ясно, что $e^{i\omega z}$ внутри Г в бесконечность не обращается и особых точек не имеет; функция $\varphi(z) = \frac{1}{1+z^2}$, как уже указывалось выше, имеет здесь единственную особую точку $z = i$ и вычет $\varphi(z)$ в этой точке равен $-\frac{i}{2}$. Отсюда вытекает, что функция $f(z) = {}^1/_2 e^{i\omega z}\varphi(z)$ в своей особой точке $z = i$ имеет вычет $-\frac{i}{4}e^{i\omega i} = -\frac{i}{4}e^{-\omega}$. В силу общих свойств аналитических функций интеграл (14), распространенный по всему огромному контуру Г, равен $\int\limits_{\sigma}\frac{e^{i\omega z}}{1+z^2}dz$, где σ — малая окружность с центром в точке $z = i$; последний же интеграл равен

$$-\frac{ie^{-\omega}}{4}\oint\limits_{\sigma}\frac{dz}{z-i} = -\frac{i}{4}e^{-\omega}2\pi i = \frac{1}{2}\pi e^{-\omega}. \qquad (17.2.15)$$

Это же значение имеет и интеграл (14), который, таким образом, оказывается не зависящим от величины R. Поэтому и интеграл (13) равен той же величине ${}^1/_2\pi e^{-\omega}$.

Аналогично получается и значение $I_- = {}^1/_2\pi e^{-\omega}$ для второго слагаемого интеграла I; таким образом, окончательно

$$I = \pi e^{-\omega}. \qquad (17.2.16)$$

Получить формулу (16) без использования функций комплексной переменной трудно. Так же трудно получить результат (16) численным интегрированием, ибо при $\omega \gg 1$ стоящее в правой части (16) (или (15)) выражение очень мало. Отдельные положительные и отрицательные полуволны периодической функции $\cos\omega x$ не будут особенно малы: их амплитуда, если говорить о волнах, отвечающих не слишком большим значениям x, имеет величину порядка 1, откуда вытекает, что вклад рассматриваемой полуволны в общее значение интеграла I имеет порядок величины $\pm\frac{1}{\omega}$ (почему?). Между тем общее значение (15) интеграла I при $\omega \gg 1$ гораздо меньше, чем $\frac{1}{\omega}$. Следовательно, положительные и отрицательные полуволны с очень большой степенью точности взаимно компенсируются, что связано с аналитичностью, т. е. с гладкостью, «функции амплитуд» $y_1 = \frac{1}{1+x^2}$.

Подобные (11) интегралы играют большую роль в теории колебаний и вообще в теоретической физике (ср. с интегралом в формуле (1.11а)). Медленно меняющийся плавный импульс $\left(\text{форма которого задается функцией времени } f(t) = \frac{1}{1+t^2}\right)$ не возбуждает высокочастотную колебательную систему потому, что при больших частотах

---

[2] Мы видим, что вдоль γ можно считать, что $|f(z)| < \frac{1}{R^2}$; отсюда и из вытекающего из самого смысла интеграла неравенства $\left|\int f(z)\,dz\right| \leqslant \int |f(z)|\,|dz|$ (см. упр. 2) с учетом того, что длина полуокружности γ растет как первая степень R, следует, что $\int\limits_{\gamma} f(z)\,dz$ с ростом R убывает, во всяком случае, не медленнее, чем величина $\frac{1}{R}$.

($\omega \gg 1$) интеграл $\int\limits_{-\infty}^{\infty} \frac{\cos \omega t}{1+t^2} dt$ будет мал. При этом лишь теория функций комплексной переменной позволяет установить, что этот интеграл с ростом частоты ω затухает экспоненциально.

Упражнения

17.2.1. Докажите прямым вычислением, что $\oint\limits_K \frac{dz}{z} = 2\pi i$, где контур интегрирования K совпадает с границей квадрата с вершинами в точках $\pm 1 \pm i$.

17.2.2. Докажите неравенство $\left| \int\limits_\Lambda f(z)\,dz \right| \leqslant \int\limits_\Lambda |f(z)|\,|dz|$, где Λ — любая дуга в комплексной плоскости.

17.2.3. Докажите, что $\int\limits_{-\infty}^{\infty} e^{-t^2/\tau^2} \cos \omega t\, dt = \tau \sqrt{\pi}\, e^{-\omega^2 \tau^2}$. (Этот результат также физически значим: так называемый гаусовый профиль e^{-x^2/τ^2} медленного импульса (см. рис. 4, *б*) меньше всего возбуждает высокочастотные колебательные системы, отвечающие большим значениям ω.)

§ 3. Аналитические функции комплексной переменной и течение жидкости

В настоящем параграфе мы расстанемся с большой темой о комплексных числах и о функциях комплексной переменной. Можно лишь удивляться и восхищаться (мы об этом уже говорили выше) тем, что введение «мнимой единицы» $i = \sqrt{-1}$ не только разрешило все трудности, возникающие при решении всевозможных квадратных уравнений, но и придало стройность всей теории алгебраических уравнений, т. е. всей алгебре, а также теории функций и математическому анализу. Естественно обобщаются на случай комплексной переменной все функции: не только многочлены и алгебраические функции (причем только здесь приобретает необходимую законченность и полноту вопрос о нулях многочленов и вопрос о степенных функциях с дробными показателями), но и экспонента, тригонометрические функции, логарифмы. Отказываясь от максимальной общности в определении функции как соответствия двух рядов величин (но ведь и в области вещественного анализа эту общность никак не удается сохранить без потерь!), мы обращаемся к понятию а н а л и т и ч е с к о й функции. Путем такого самоограничения удается колоссально повысить информативность теории — на аналитические функции комплексной переменной полностью переносится весь аппарат высшей математики (т. е. понятия производной, интеграла, дифференциального уравнения, степенного ряда); для них можно также высказать много утверждений, несправедливых, не имеющих места для функций вообще. В частности, выявляется определяющая роль точек (значений переменных), в которых функция уходит в бесконечность, устанавливается, что поведение функции в этих точках определяет широкий круг свойств функции, например позволяет без вычислений указать значения интегралов от функции по любым контурам.

Доказывается, что в принципе можно задать функцию комплексной переменной на сколь угодно малом отрезке — и это ее полностью определит на всей плоскости, во всей области изменения переменной. Выясняется, что аналитичность определяется не тем, есть ли простая удобная формула, выражающая искомую функцию через переменную z: аналитической функцией может быть, скажем, и и н т е г р а л, не берущийся в элементарных функциях $\left(\text{типа интеграла } \int\limits_a^z e^{-z^2/2} dz\right)$, и р е ш е н и е д и ф ф е р е н ц и а л ь н о г о у р а в н е н и я, даже если это уравнение в традиционном смысле «неразрешимо», т. е. не позволяет выписать зависимость неизвестной функции от аргумента в виде явной формулы, и зависимость, задаваемая в виде б е с к о н е ч н о г о р я д а. Важно лишь, чтобы значение $w = f(z)$ функции зависело именно от $z = x + iy$, а не от x и y в отдельности, чтобы величина w не разрушала простую связь между x и y в их комбинации $x + iy$.

Функции комплексной переменной оказываются мощным рычагом для решения математических, физических и технических задач. В этих задачах исходные данные, естественно, задаются с помощью вещественных величин.

Ответ тоже по смыслу задачи должен быть вещественным, — и если С. Маршака испугали «два землекопа и две трети», то мнимый землекоп — это, пожалуй, еще ужаснее. И тем не менее часто оказывается полезным вводить на промежуточной стадии решения задачи, между ее формулировкой и нахождением ответа, комплексные величины, рассматривая (правда, только временно!)

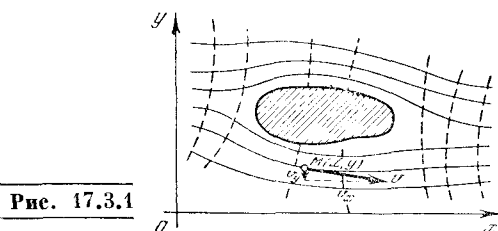


Рис. 17.3.1

именно «комплексного землекопа» вместо вещественного (ср. с § 1).

Так, например, мы видели (см. § 2), что для вычисления $\int\limits_{-\infty}^{\infty} \frac{\cos \omega x}{1+x^2}\,dx$ (чисто вещественный интеграл!) уместно записать $\cos \omega x$ как $^1/_2 e^{i\omega x} + {}^1/_2 e^{-i\omega x}$, введя тем самым в задачу, казалось бы, совершенно неуместные здесь «мнимости». При этом если нам даже и неизвестны конкретные формулы, выражающие какой-то закон природы, то часто одно предположение о существовании выражающей этот закон аналитической функции комплексной переменной (какой угодно, хотя бы не выражающейся формулой) уже оказывается информативным и плодотворным. Из самого факта существования такой функции сразу же возникают определенные соотношения, порожденные аналитичностью этой (неизвестной нам!) функции. Такой прием широко используется на переднем крае современной теоретической физики; с ним связаны, в частности, так называемые *дисперсионные соотношения*, подробно говорить о которых здесь, к сожалению, неуместно. Самое большое чудо — это тот факт, что мнимая единица и комплексные величины необходимым образом фигурируют в квантовой теории — в физической теории микромира. При этом место «мнимостей» в квантовой теории настолько значительно, что сегодня, подчас, при изложении элементарной аналитической геометрии будущим физикам и химикам считают нужным чуть ли не с самого начала говорить о «комплексных» плоскости и пространстве, где точки имеют комплексные координаты, хотя наглядно представить себе такое пространство, понятно, невозможно [1].

Видимо, исторически первыми серьезными приложениями теории аналитических функций комплексной переменной, с которыми частично было связано даже зарождение самой этой теории (см. с. 432), явились приложения ее к *гидро-* и *аэромеханике*, к учению о движении жидкостей и газов. Пусть мы имеем *плоское* (*плоскопараллельное*) течение (несжимаемой) жидкости постоянной плотности, при котором вектор скорости **v** любой частицы жидкости в любой момент времени параллелен (горизонтальной) (x, y)-плоскости и весь столб жидкости, отвечающий данным координатам x и y и разным высотам z, движется одинаково; в таком случае мы можем забыть о (трехмерном) пространстве и ограничиться рассмотрением движения в плоскости x, y. Пусть скорость **v** (вектор!) течения в точке $M=M(x, y)$ имеет компоненты v_x и v_y (рис. 1); ясно, что $v_x=v_x(x, y)$ и $v_y=v_y(x, y)$ зависят от точки M, т. е. являются функциями этой точки — функциями координат[2] x и y.

Во многих случаях допустимым является предположение о существовании двух таких функций точки $\varphi(x, y)$ и $\psi(x, y)$, что [3]

$$v_x = -\frac{\partial\varphi(x, y)}{\partial x}, \quad v_y = -\frac{\partial\varphi(x, y)}{\partial y}, \tag{17.3.1}$$

$$v_x = -\frac{\partial\psi(x, y)}{\partial y}, \quad v_y = \frac{\partial\psi(x, y)}{\partial x}. \tag{17.3.1a}$$

---

[1] См., например, элементарный учебник: *Головина Л. И.* Линейная алгебра и некоторые ее приложения. 3-е изд. М.: Наука, 1979.

[2] Мы здесь для простоты ограничиваемся так называемым *стационарным* течением жидкости, при котором скорость не меняется с течением времени и функции $v_x(x, y)$ и $v_y(x, y)$ от времени t не зависят.

[3] Так как функции φ и ψ определяются лишь значениями (1), (1а) своих производных, то обе они имеют характер потенциалов, т. е. определены лишь с точностью до произвольных слагаемых C_1 и C_2, позволяющих произвольно выбрать точку нулевых значений φ и ψ; при этом физический смысл имеют лишь разности значений φ (или ψ) в двух точках, но не сами эти значения. (Заметим еще, что поскольку величины v_x и v_y, разу-

Функцию φ называют *потенциалом скоростей*, а движение жидкости, характеризующееся наличием потенциала скоростей, — *потенциальным движением*. Можно сказать, что движение жидкости в окрестности данной точки $M=M(x, y)$ тогда и только тогда потенциально, когда оно сводится к переносу массы жидкости в направлении $\mathbf{v}$ и (допустимой в силу текучести жидкости) перестройке (деформации) маленького ее объема, но не включает вращения всей жидкости вокруг M, представление о котором могут дать наблюдающиеся иногда на реках водовороты или маленькие «искусственные водовороты» (воронки), создаваемые веслами при гребле. В соответствии с этим потенциальное движение жидкости называется также *безвихревым*, а (вообще говоря, изолированные) точки, в которых нарушается условие потенциальности движения, именуются *вихревыми точками* (или просто *вихрями*). Разность

$$\frac{\partial v_y}{\partial x}-\frac{\partial v_x}{\partial y}, \tag{17.3.2}$$

равная нулю в случае потенциального движения жидкости (почему?) и, следовательно, отличная от нуля только в вихревых точках, характеризует *интенсивность* вихря.

Функцию $\psi=\psi(x, y)$ называют *функцией тока* жидкости. Можно доказать, что наличие в точке $M(x, y)$ функции тока ψ равносильно предположению о том, что в единицу времени внутрь любого маленького контура, окружающего точку M, втекает ровно столько же (несжимаемой) жидкости, сколько и вытекает из этого контура. Таким образом, наличие функции тока ψ в данной точке $M=M(x, y)$ обеспечивает отсутствие в точке M источника или стока жидкости. При этом сумма

$$\frac{\partial v_x}{\partial x}+\frac{\partial v_y}{\partial y} \tag{17.3.2a}$$

(обязательно равная нулю в случае наличия в точке $M=M(x, y)$ функции тока ψ) характеризует *интенсивность* источника или стока, а знак этой суммы — характер «особой точки» M: если сумма (2а) положительна, то в точке M мы имеем *источник* (в окружающий M контур втекает меньше жидкости, чем вытекает из него), а если она отрицательна — *сток*. (Знак разности (2) характеризует *направление вращения* жидкости в вихревой точке.)

Нетрудно видеть (см. упр. 1), что линии $\psi=\text{const}$ представляют собой не что иное, как *линии тока*, по которым движутся частицы жидкости: вдоль линии тока (характеризуемой тем, что касательная к такой линии в каждой ее точке M имеет направление вектора скорости $\mathbf{v}$) величина функции тока ψ остается постоянной. *Эквипотенциальные линии*, характеризуемые условием $\varphi(x, y)=\text{const}$, напротив, перпендикулярны в каждой своей точке вектору скорости (см. упр. 2): это суть условные поверхности, через которые движется жидкость (причем в сторону убывания потенциала). (На рис. 1 эквипотенциальные линии изображены пунктиром.) Таким образом, сеть линий $\varphi=\text{const}$ и $\psi=\text{const}$ дает достаточно полное представление о движении жидкости в целом.

Предположим теперь, что в окрестности данной точки $M(x, y)$ в жидкости отсутствуют как вихревые точки, так и источники и стоки, другими словами, что здесь имеются обе функции φ и ψ. Очевидно, что из (1) и (1а) следует:

$$-v_x=\frac{\partial\varphi}{\partial x}=\frac{\partial\psi}{\partial y},\quad -v_y=\frac{\partial\varphi}{\partial y}=-\frac{\partial\psi}{\partial x}. \tag{17.3.3}$$

Но уравнения (3) близки по форме уравнениям Коши—Римана (15.2.6), связывающим вещественную часть и коэффициент при мнимой части аналитической функции $w=w(z)=u(x, y)+iv(x, y)$ комплексной переменной $z=x+iy$. Таким образом, если отождествить плоскость x, y с плоскостью комплексной переменной $z=x+iy$, то мы, естественно, придем к аналитической функции

$$w=w(z)=\varphi+i\psi \tag{17.3.4}$$

переменной z, теснейшим образом связанной с течением жидкости.

Функция комплексной переменной (4) называется *комплексным потенциалом* течения жидкости. Комплексный потенциал полностью характеризует те-

мется, имеют размерности скорости (м/с), а координаты x и y — размерности длины (м), то определяемые по (1) и (1а) функции φ и ψ имеют размерность м^2/с.)

чение, так что изучение всевозможных (плоских) течений сводится к изучению аналитических функций $w(z)$. Производная комплексного потенциала

$$v=\frac{dw}{dz}\left(=\frac{\partial\varphi}{\partial x}+i\frac{\partial\psi}{\partial x}=-v_x+iv_y\right) \tag{17.3.5}$$

(ср. (15.2.3)) также является аналитической функцией комплексной переменной z. Эта функция называется *комплексной скоростью* течения. Последнее название подчеркивает связь комплексного числа v и вектора $\mathbf{v}$; очевидно,

$$|v|^2=(-v_x)^2+(v_y)^2=v_x^2+v_y^2=|\mathbf{v}|^2,$$

т. е. модуль (абсолютная величина) числа v равен численной величине скорости $\mathbf{v}$; с другой стороны, $\operatorname{Arg} v$ характеризует направление вектора $\mathbf{v}$ (см. упр. 3). Течения, задаваемые комплексными потенциалами $w=\varphi+i\psi$ и $iw=-\psi+i\varphi$, можно назвать *сопряженными*; они характеризуются тем, что линии тока одного из них совпадают с эквипотенциальными линиями второго, и наоборот.

Весьма близко к намеченной схеме и использование аналитических функций в *теории электромагнитного поля*[4]. Рассмотрим плоское электрическое поле, причем будем считать, что в рассматриваемой (плоской) области отсутствуют свободные электрические заряды и переменные магнитные поля (в силу принципа индукции равносильные наличию зарядов). Пусть $\Phi=\Phi(x, y)$ — *потенциал* электрического поля в точке $M=M(x, y)$; тогда *напряженность* поля в этой точке задается вектором напряженности $\mathbf{E}=(E_x, E_y)$ с координатами

$$E_x=-\frac{\partial\Phi}{\partial x}, \quad E_y=-\frac{\partial\Phi}{\partial y}. \tag{17.3.6}$$

Наряду с потенциалом Φ можно ввести в рассмотрение и *функцию тока поля* $\Psi=\Psi(x, y)$ такую, что[5]

$$E_x=-\frac{\partial\Psi}{\partial y}, \quad E_y=\frac{\partial\Psi}{\partial x}. \tag{17.3.6a}$$

При этом линии $\Psi=\text{const}$ суть *силовые линии* нашего электрического поля, а линии $\Phi=\text{const}$ — его *эквипотенциальные линии*. В силу (6) и (6а) функция

$$W=\Phi+i\Psi \tag{17.3.7}$$

является аналитической функцией комплексной переменной $z=x+iy$; она называется *комплексным потенциалом* поля. Производную

$$E=\frac{dW}{dz}\left(=\frac{\partial\Phi}{\partial x}+i\frac{\partial\Psi}{\partial x}=-E_x+iE_y\right) \tag{17.3.8}$$

комплексного потенциала можно назвать *комплексной напряженностью* поля, поскольку это комплексное число очень близко к вектору напряженности $\mathbf{E}$; в частности, очевидно, $|E|=|\mathbf{E}|$. Электрические поля с комплексными потенциалами $W=\Phi+i\Psi$ и $iW=-\Psi+i\Phi$ иногда называют *сопряженными*; линии тока одного из них совпадают с эквипотенциальными линиями второго, и наоборот.

Мы не остановимся здесь подробнее на дальнейших положениях математической теории течения жидкости и теории электрического поля, имеющей дело с функциями двух переменных x и y и соответственно этому с дифференциальными уравнениями, в которых фигурируют частные производные этих функций, — с так называемыми *дифференциальными уравнениями в частных производных* (ср. с § 10.8). Укажем лишь, что теория аналитических функций комплексной переменной играет в этих чисто прикладных вопросах очень большую роль.

Упражнения

17.3.1. Докажите, что: а) линия $\psi=\text{const}$ является линией тока жидкости; другими словами, касательная к этой линии в каждой ее точке $M(x, y)$ имеет направление вектора $\mathbf{v}(x, y)$; б) если $\smile AB$ — произвольная дуга (в плоскости течения жидкости) с концами $A=A(x_1, y_1)$, $B=B(x_2, y_2)$, то поток жидкости через эту дугу в единицу времени равен разности $\psi(x_2, y_2)-\psi(x_1, y_1)$ (приращению функции тока вдоль дуги AB).

---

[4] Точнее, здесь речь идет лишь об *электро-* и *магнитостатических* явлениях, поскольку в общих задачах электромагнетизма не удается обойтись всего двумя координатами x, y (вместо четырех: x, y, z и t, где t — это время).

[5] Поскольку потенциал Φ (а значит, и функция тока Ψ) имеет размерность В, а координаты x и y имеют размерность м (размерность длины), то ясно, что компоненты E_x и E_y вектора напряженности электрического поля имеют размерность В/м (можно говорить, что В/м — размерность вектора $\mathbf{E}$).

17.3.2. Докажите, что эквипотенциальные линии $\varphi=\text{const}$ и линии тока $\psi=\text{const}$ взаимно перпендикулярны в каждой точке потока жидкости (точнее, касательные в точке $M(x, y)$ к проходящим через эту точку линии тока и эквипотенциальной линии взаимно перпендикулярны).

17.3.3. Пусть $V=V(x, y)$ — такое комплексное число, что $\overrightarrow{OV}=\mathbf{v}$ (где $\overrightarrow{OV}$ — это вектор с началом $O(0)$ и концом V, а $\mathbf{v}$ — вектор скорости в точке $M(x, y)$). Докажите, что точки v и V комплексной плоскости, где v — комплексная скорость течения, симметричны относительно мнимой оси (и, следовательно, $\operatorname{Arg} v=\pi-\operatorname{Arg} V$).

17.3.4. Опишите (плоское) движение (несжимаемой) жидкости, задаваемое следующими выражениями комплексного потенциала $w=\varphi+i\psi$: а) $w=az$, где a — вещественное или чисто мнимое число; б) $w=az^2$ (a — вещественное или чисто мнимое число); в) $w=\frac{a}{z}$ (те же условия для a); г) $w=\ln z$ и $w=i\ln z$.

§ 4. Применения дельта-функции

Если § 1—3 настоящей главы примыкали к главам 14 и 15, т. е. в них мы имели дело с комплексными числами и аналитическими функциями комплексной переменной, то настоящий параграф продолжает тему гл. 16, которую необходимо перед его чтением снова просмотреть. Укажем еще, что первый пример, с разбора которого мы начинаем этот параграф, тесно связан с содержанием § 9.12, а второй пример, касающийся уравнения Ньютона (1), продолжает изложенные в § 9.5 рассуждения об импульсе силы и о движении частицы под действием кратковременного импульса, например под влиянием удара; поэтому целесообразно перед разбором этих примеров вспомнить соответствующие места гл. 9.

Покажем прежде всего, как дельта-функция позволяет сократить и сделать более удобной запись условий во многих задачах.

Рассмотрим стержень переменного сечения, к которому прикреплено несколько отдельных точечных грузов m_1, m_2 и т. д. (рис. 1). Пусть масса, приходящаяся на единицу длины стержня, выражается функцией $\sigma(x)$. Масса всего стержня без грузов равна $\int_a^b \sigma(x)\,dx$, а вместе с грузами она равна

$$M=\int_a^b \sigma(x)\,dx+\sum_i m_i,$$

где суммирование распространяется по всем приложенным к стержню грузам. Координата центра тяжести задается следующим выражением:

$$X=\frac{1}{M}\left(\int_a^b x\sigma(x)\,dx+\sum_i x_i m_i\right).$$

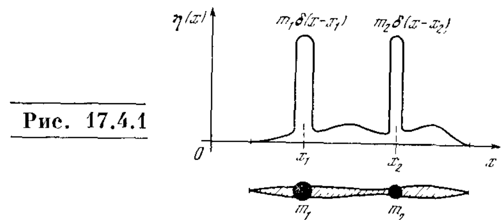


Рис. 17.4.1

Момент инерции относительно начала координат будет таков:

$$I=\int_a^b x^2\sigma(x)\,dx+\sum x_i^2 m_i.$$

С помощью дельта-функции можно включить отдельные массы в обобщенную функцию плотности. Обозначим новую функцию $\eta(x)$. Она выражается формулой

$$\eta(x)=\sigma(x)+\sum_i m_i\delta(x-x_i).$$

В самом деле, рассматривая общее распределение массы по стержню, можно сказать, что в тех точках, где находятся грузы, плотность имеет бесконечные подскоки. С помощью функции $\eta(x)$ все величины записываются единообразно и более кратко:

$$M=\int_a^b \eta(x)\,dx,\qquad X=\frac{1}{M}\int_a^b x\eta(x)\,dx,$$

$$I=\int_a^b x^2\eta(x)\,dx.$$

Понятие дельта-функции позволяет объединить непрерывно распределенные и точечные массы в одном общем выражении.

Другой пример применения дельта-функции относится к движению мате-

риальной точки. Основное уравнение Ньютона (второй закон Ньютона) имеет вид

$$m\frac{d^2x}{dt^2}=F(t) \qquad (17.4.1)$$

(см. § 9.4). Вспомните изложенные в § 9.5 соображения о том, что действие импульса силы не зависит от закона изменения силы, если только сила достаточно кратковременна. Эти соображения аналогичны рассуждениям § 16.4 о том, что дельта-функцию можно построить из р а з л и ч н ы х функций $\varphi(x)$, и об условиях, когда можно

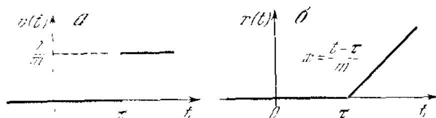


Рис. 17.4.2

обычную функцию $\psi(x)$ заменить обобщенной, особенной функцией $\delta(x)$.

Если конкретная форма функции силы несущественна в задаче об ударе, это значит, что $F(t)$ можно заменить на дельта-функцию, $F(t)\to I\delta(t-\tau)$, где τ — момент удара, а $I=\int F(t)\,dt$ — импульс силы. Проведем формально по всем правилам интегрирование уравнения движения под действием единичной дельта-силы. Пусть до удара частица покоилась в начале координат: $x=0$, $v=\frac{dx}{dt}=0$ при $t=-\infty$. Примем, что (разумеется, малосущественный) множитель I в выражении для «обобщенной силы» равен 1; тогда уравнение (1) примет вид

$$m\frac{d^2x}{dt^2}=m\frac{dv}{dt}=\delta(t-\tau). \qquad (17.4.2)$$

Интегрируя (2) и учитывая, что $\frac{d^2x}{dt^2}=\frac{dv}{dt}$, где $v(t)$ — это скорость, получим

$$v(t)=\frac{1}{m}\int_{-\infty}^{t}\delta(t-\tau)\,dt=\frac{1}{m}\theta(t-\tau), \qquad (17.4.3)$$

где функция $\theta(x)$ выражается формулой (16.3.1) (см. ее график на рис. 16.3.1). Таким образом, скорость выражается ступенчатой функцией времени (рис. 2, *а*): $v=0$ при $t<\tau$ и $v=\frac{1}{m}$ при $t>\tau$.

Следующий шаг заключается в определении пути. Так как $v=\frac{dx}{dt}$, то нам достаточно проинтегрировать обе части (3); мы получим

$$x=\begin{cases}0 & \text{при } t<\tau,\\ \frac{1}{m}(t-\tau) & \text{при } t>\tau\end{cases} \qquad (17.4.4)$$

(почему?). График пути изображен на рис. 2, *б*.

Для графика пути $x=x(t)$ характерен излом в точке $t=\tau$. Здесь еще раз мы убеждаемся в том, что вторая производная функции, имеющей излом, содержит дельта-функцию: функция $x(t)$ имеет излом; согласно уравнению движения сила пропорциональна $\frac{d^2x}{dt^2}$; выражение для $x(t)$ в виде «функции с изломом» как раз и получено при силе, пропорциональной $\delta(t-\tau)$, так что при наличии излома формула для $\frac{d^2x}{dt^2}$ содержит δ-функцию.

Теперь сделаем следующий шаг. Задача о движении тела под действием заданной силы л и н е й н а. Это значит, что если есть два решения $x_1(t)$ и $x_2(t)$, отвечающие действиям двух разных сил $F_1(t)$ и $F_2(t)$, то сумма решений $x(t)=x_1(t)+x_2(t)$ является решением, соответствующим действию суммы сил $F(t)=F_1(t)+F_2(t)$. Это свойство (*принцип суперпозиции*) есть следствие того простого факта, что вторая производная суммы функций есть сумма вторых производных каждой функции:

$$\frac{d^2x}{dt^2}=\frac{d^2(x_1+x_2)}{dt^2}=\frac{d^2x_1}{dt^2}+\frac{d^2x_2}{dt^2}.$$

Принимая во внимание, что $\frac{d^2x_1}{dt^2}=\frac{F_1(t)}{m}$, $\frac{d^2x_2}{dt^2}=\frac{F_2(t)}{m}$, получим

$$\frac{d^2x}{dt^2}=\frac{F_1(t)}{m}+\frac{F_2(t)}{m}=\frac{F(t)}{m},$$

что и требовалось — сумма решений x_1+x_2 описывает движение под действием суммы сил.

Нужно сделать только одну оговорку: решение уравнения движения зависит не только от закона изменения силы, но и от начальных условий, т. е. от начального положения и начальной скорости рассматриваемой массы. Если мы выберем эти условия так, что $x_1=0$,

$\frac{dx_1}{dt}=0$ при $t=-\infty$, а также $x_2=0$, $\frac{dx_2}{dt}=0$ при $t=-\infty$, то и сумма решений x будет удовлетворять тому же условию: $x=0$, $\frac{dx}{dt}=0$ при $t=-\infty$.

Теперь объединим соображения линейности и известное решение для дельта-функции, с тем чтобы получить общее решение уравнения для силы, произвольно зависящей от времени. Разобьем график силы $F(t)$ на полоски шириной $\Delta\tau$ (рис. 3). Что собой представляет отдельная полоска, расположенная между значениями τ и $\tau+\Delta\tau$ времени t? Мы изменим здесь обозначения, чтобы оставить букву t для «текущего» времени, меняющегося от $-\infty$ до ∞, тогда как τ относится к данной избранной полоске. Итак, высота полоски $F(\tau)$, ширина $\Delta\tau$; поэтому площадь (т. е. импульс силы) равна $F(\tau)\,\Delta\tau$. Так как полоска отвечает значению $t=\tau$, то, очевидно, ее можно заменить дельта-функцией с коэффициентом, равным импульсу: $F(\tau)\,\Delta\tau\delta(t-\tau)$. Решение уравнения движения для дельта-функции мы уже знаем. Обозначим его $x_1(t, \tau)$. Решение как функция времени t зависит от момента приложения силы τ. Напоминаем, что

$$x_1(t, \tau)=\begin{cases}0 & \text{при } t<\tau,\\ \dfrac{t-\tau}{m} & \text{при } t>\tau.\end{cases} \qquad (17.4.4а)$$

Одна из полосок, на которые разложена сила, простирающаяся от τ до $\tau+\Delta\tau$, представляет собой функцию $\delta(t-\tau)$, взятую с коэффициентом $F(\tau)\,\Delta\tau$. Благодаря линейности уравнения решение для силы в виде такой полоски получится умножением x_1 на тот же коэффициент: $F(\tau)\,\Delta\tau\cdot x_1(t, \tau)$. Это есть решение, относящееся к действию одной полоски.

Теперь используем линейность и выпишем решение для функции $F(t)$, которую мы рассматриваем как сумму полосок. Ясно, что суммирование здесь следует заменить интегрированием. Итак,

$$x(t)=\int_{-\infty}^{\infty} x_1(t, \tau)\,F(\tau)\,d\tau. \qquad (17.4.5)$$

На первый взгляд написанная формула имеет странный вид: координата x в момент t выражена интегралом по τ от $-\infty$ до ∞, т. е. в выражение входит сила во в с е моменты времени. Между тем ясно, что вид силы после момента t не влияет на предыдущее движение. Однако в выражении (5) для $x(t)$ нет никакой ошибки. Свойства функции $x_1(t, \tau)$ обеспечивают разумные свойства решения. В самом деле, $x_1(t, \tau)=0$ при $t<\tau$. Следовательно, при интегрировании по τ в действительности не нужно брать $\tau>t$, так как там тождественно равна нулю подынтегральная функция за счет обращения

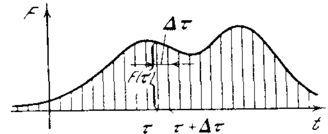


Рис. 17.4.3

в нуль множителя $x_1(t, \tau)$. Вспоминая формулу (4а) для $x_1(t, \tau)$, получим

$$x(t)=\frac{1}{m}\int_{-\infty}^{t}(t-\tau)\,F(\tau)\,d\tau. \qquad (17.4.5а)$$

Такой способ получения решения имеет очень большое общее значение. Вывод: если для линейной системы известно решение, относящееся к воздействию дельта-функции, то решение, относящееся к воздействию произвольной функции ($F(t)$ в нашем примере), получается простым суммированием (точнее, интегрированием).

Идеи линейности и сложения (*суперпозиции*) решений относятся не только к таким простым задачам, как движение точки; эти идеи справедливы для огромных областей математики, физики, естествознания. Бывает и так, что система сложна и решить уравнения нельзя даже для самого простого воздействия дельта-функцией. Решение, соответствующее дельта-функции, иногда можно получить опытным путем. В других случаях такое решение можно получить из физических соображений (см. упр. 1). Дальше вступают в силу соображения линейности — и мы получаем ответ для л ю б о й воздействующей функции. Решение, соответствующее дельта-функции (т. е. $x_1(t, \tau)$ в примере выше), столь важно, что оно имеет специальное название — ф у н к ц и я Г р и н а. Забавно то, что анг-

лийский математик Грин [1], именем которого названа функция, жил в XIX в. и, естественно, не знал о дельта-функции. Но только введение дельта-функ-

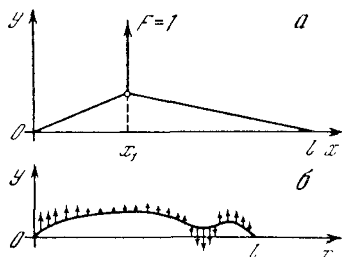


Рис. 17.4.4

ции позволило ясно и кратко объяснить суть функции Грина.

Таких примеров в математике немало: ведь и до изобретения производных и интегралов было известно много результатов, относящихся к касательным, площадям и объемам. Движение науки заключается не только в завоевании новых высот, новых результатов, но и в популяризации и упрощении выводов, полученных ранее. Задачей книги, чтение которой Вы сейчас заканчиваете, как раз и было облегчение понимания классического наследия — основ высшей математики.

[1] Автор замечательных работ по математике и математической физике Джордж Грин (1793—1841) происходил из очень бедной семьи; он не получил никакого образования и сравнительно поздно самостоятельно познакомился с математикой и физикой. Впрочем, в конце жизни, главным образом благодаря поддержке влиятельного У. Томпсона (лорда Кельвина), Грин был приглашен профессором математики в Кембриджский университет.

Упражнения

17.4.1. Рассмотрим струну, натянутую силой k, с концами, закрепленными в точках $x=0$, $x=l$. Считая отклонение малым, определите по закону параллелограмма сил форму струны при действии единичной нагрузки в точке $x=x_1$ (рис. 4, *а*). Получите формулу отклонения струны под действием силы, распределенной по ее длине по произвольному закону $f(x)$ (рис. 4, *б*).

17.4.2. Найдите движение маятника под действием силы, выражающейся дельта-функцией, т. е. решите уравнение $m\frac{d^2x}{dt^2}=-kx+\delta(t-\tau)$ при начальных условиях: $x=0$, $\frac{dx}{dt}=0$ при $t=-\infty$. С помощью этого решения найдите движение маятника под действием силы, зависящей от времени по произвольному закону.

ЗАКЛЮЧЕНИЕ. ЧТО ЖЕ ДАЛЬШЕ?

Высшая математика, а точнее, математический анализ, или дифференциальное и интегральное исчисление, дает возможность решить многие задачи, не поддающиеся решению методами элементарной математики: арифметики, алгебры, геометрии. Огромное значение имеет формирование новых понятий, таких, как мгновенная скорость, ускорение, импульс силы; эти понятия (и многие другие, из других областей знания) определяются точно только на языке производных и интегралов.

То, что Вы узнали из настоящей книги, — это только малая часть всей математической науки, малая часть из тех разделов математики, которые применяются в технике, физике, естествознании. В предисловии мы говорили о том, что культурный человек независимо от специальности должен иметь общие представления о математике. Первые три части нашей книги дают именно такие общие сведения. Но если Вы будете специализироваться в технике, химии и особенно в физике, то этих общих сведений будет Вам недостаточно. Здесь мы хотим коротко охарактеризовать те области физики и связанные с ними разделы математики, которые Вам предстоит изучать в дальнейшем. Необходимо, однако, предварительно предупредить о следующем.

Части 1—3 книги были написаны как учебник, так что, приложив достаточно внимания и труда, можно было полностью, во всех деталях понять написанное. Имейте в виду, что ниже, рассказывая о трудных вопросах, мы вынуждены будем отступить от стиля учебника. На небольшом числе страниц нельзя, разумеется, изложить содержание сложных математических теорий; можно только попытаться дать общее представление о них, с тем чтобы заинтересовать читателя. Некоторые из названных в списке литературы книг помогут Вам расширить представление о математике и математической физике.

Попробуем охарактеризовать вкратце то общее, что объединяло все задачи, которыми мы занимались выше. Это были задачи о движении одной частицы в механике или о токе в цепи. Мы имели дело с функциями одной переменной — времени; число рассматриваемых функций могло равняться единице (сила тока как функция времени) или двум (скажем, положение тела $x(t)$ и скорость $v(t)$).

Напрашивается естественное чисто количественное обобщение: дальше можно рассмотреть задачи о движении двух, трех тел и т. д. За этим «и т. д.» в далекой перспективе видна и задача о движении газа или жидкости, о движении молекул газа. Но ведь 1 г водорода — это примерно $3 \cdot 10^{23}$ молекул, т. е. $3 \cdot 10^{23}$ отдельных тел. Ясно, что надежды обойтись здесь старыми методами не остается — тут необходимы какие-то совершенно новые подходы. Ведь не только решить эти $3 \cdot 10^{23}$ уравнений нельзя, но и просто выписать их все не хватит ни времени, ни бумаги. Да мы и не можем выписать все эти уравнения: ведь ни точное число молекул, ни их начальное расположение, определяющее начальные условия рассматриваемых уравнений, никогда не будут нам известны.

Здесь помогают новые науки, возникшие гораздо позже дифференциального и интегрального исчисления (по существу — лишь в XIX в.), — гидродинамика и газодинамика (аэродинамика). Эти науки существенно отличаются от механики точки и постановкой задач, и способами их решения (ср. § 9.14, 9.15 и 17.3). Мы совсем и не интересуемся точным числом молекул газа, заключенным в том или ином фиксированном его объеме. Вместо этого в рассмотрение вводится плотность распределения газа в пространстве $\rho(x, y, z)$ (т. е. масса газа, приходящаяся на единицу объема [1]), давление

[1] Собственно, само введение плотности в точке $\rho(x, y, z)$ требует операции, подобной дифференцированию. Эта величина определяется как предел средней плотности в малом объеме (частное массы газа и объема) при стремлении рассматриваемого объема к нулю. Изменения плотности, зависящие от того, что одна или несколько молекул случайно войдут в какой-то объем или вылетят из него, нас не интересуют.

газа $p(x, y, z)$, скорость газа в различных точках пространства. Надо добавить, что все эти величины зависят и от времени t, так что уместно например, писать $p=p(x, y, z, t)$.

Таким образом, от задачи о нескольких функциях одной переменной мы переходим к задаче о нескольких функциях нескольких независимых переменных.

Соответственно при составлении уравнений, описывающих движение газа и другие его характеристики, появляются *частные производные* по времени по пространственным координатам, например $\frac{\partial\rho}{\partial t}$ и $\frac{\partial\rho}{\partial x}$, $\frac{\partial\rho}{\partial y}$, $\frac{\partial\rho}{\partial z}$, где, скажем,

$$\frac{\partial\rho}{\partial y}=\lim_{\Delta y\to 0}\frac{\rho(x, y+\Delta y, z, t)-\rho(x, y, z, t)}{\Delta y}.$$

Важнейший раздел математической физики составляет именно исследование *уравнений в частных производных*, о которых мы лишь весьма бегло упомянули в § 10.8. Уравнения в частных производных описывают движение жидкостей, газов и твердых тел, распространение тепла в этих средах, явления диффузии атомов и молекул. При этом значение уравнений в частных производных для физики настолько велико, что эти уравнения зачастую называют *уравнениями математической физики*.

Заметим еще, что скорость газа $\mathbf{v}=\mathbf{v}(x, y, z)$ (или $\mathbf{v}(x, y, z, t)$) есть векторная величина, т. е. в каждой точке заданы значение и направление скорости. Иными словами, можно сказать, что заданы три составляющие v_x, v_y, v_z вектора $\mathbf{v}$, зависящие от точки в пространстве и от времени; это обстоятельство приносит новые усложнения (или упрощения) в стоящие в гидро- и аэромеханике задачи, которых мы здесь не коснемся.

В перечисленных выше теориях хотя бы в принципе можно было бы по-прежнему рассматривать отдельные частицы и набор большого числа «обыкновенных» функций, т. е. функций одной переменной — времени. Но существуют физические теории, где такой подход никак невозможен. К их числу прежде всего принадлежит *теория электромагнетизма*.

Пусть рассматриваются два покоящихся точечных заряда. Сила, действующая между ними, зависит от их положения (от расстояния r между ними); казалось бы, это задача о шести функциях (координатах x_1, y_1, z_1 первого заряда и координатах x_2, y_2, z_2 второго заряда), зависящих от одной переменной — времени. Движение зарядов в первом приближении мало меняет дело: надо только учитывать, что возникает еще магнитное взаимодействие между зарядами, зависящее от их скоростей.

Важнейший факт, требующий принципиально нового подхода, заключается в том, что существует запаздывание взаимодействия, передача взаимодействия со скоростью света. Действие одного заряда на другой в момент t зависит от положения (и скорости) первого заряда в некий предыдущий момент времени $t-\tau$, отстоящий от рассматриваемого момента t на конечное время запаздывания $\tau=\frac{r}{c}$, где c — скорость света. Что же происходит в течение этого промежутка времени?

В пространстве между зарядами, в пустоте, существуют электрическое и магнитное поля. В каждой точке пространства имеются два вектора $\mathbf{E}$ и $\mathbf{H}$. Величина и направление этих векторов зависят от наличия зарядов и от их движения. Малые «пробные» заряды, помещенные в ту или иную точку пространства, дают возможность измерить поля $\mathbf{E}$ и $\mathbf{H}$ в данной точке. Но теория электромагнетизма этим не исчерпывается! В пустом пространстве, в котором нет никаких зарядов, электрическое и магнитное поля действуют друг на друга: изменение (в зависимости от времени) одного поля связано уравнениями Максвелла с производными другого поля по координатам. Производная магнитного поля по времени создает кольцевое электрическое поле, производная электрического поля играет ту же роль, что и электрический ток, и создает магнитное поле. Получается стройная картина: заряды создают поля в той точке, где они находятся, а взаимодействие полей между собой переносит информацию о расположении и движении зарядов по всему пространству.

Решающую роль в развитии теории имеет рассмотрение двух векторов $\mathbf{E}$ и $\mathbf{H}$ или, другими словами, шести вели-

чин E_x, E_y, E_z, H_x, H_y, H_z как функций четырех величин: трех координат x, y, z и времени t. Но задание таких функций в определенный момент времени — это задание бесчисленного количества их значений во всех точках пространства. Теория труднее, но и богаче результатами по сравнению с теорией движения одной или нескольких частиц.

Математическая теория электромагнитного поля — это теория уравнений в частных производных, похожая на теорию упругости, акустику или газодинамику. Разница лишь в том, что в последнем случае уравнения получаются путем идеализации: о плотности газа мы говорим, отвлекаясь от отдельных молекул, — и только в этом приближенном смысле газ можно рассматривать как непрерывную среду, характеризуемую непрерывной функцией $\rho(x, y, z, t)$ (плотностью газа). Электрическое (точнее, электромагнитное) поле фактически является непрерывной функцией пространственных переменных и времени.

Гидродинамика, начала которой были заложены еще в XVIII в. Д. Бернулли, Д'Аламбером и Эйлером, подготовила математический аппарат для теории электромагнетизма (разработанной во второй половине XIX в. Джеймсом Клерком Максвеллом (1831—1879)). Не удивительно поэтому, что вначале пытались перенести в электромагнитную теорию идеи механики — думали, что есть особое вещество, *эфир*, заполняющий всю среду, и что движение этого вещества ответственно за электрические и магнитные явления. Мы знаем, что математическая аналогия между гидродинамикой и электромагнетизмом, близость соответствующих уравнений остались, а физический смысл электромагнитной теории оказался другим, не сводящимся к механике, — после длительных и напряженных дискуссий ученым пришлось полностью отказаться от представлений об эфире.

Говоря о математической теории, обязательно нужно сказать не только о постановке задачи и исходных уравнениях, но и о характере результатов.

Для уравнений в частных производных можно выделить два типа решений. Один тип, характерный для явлений, развивающихся в ограниченном объеме, — это собственные колебания с определенными частотами. Тело данной формы имеет некий набор частот.

Вспомните маятник: у него есть определенная частота колебаний. При воздействии (периодической) внешней силы F на маятник, когда частота F близка к (собственной) частоте колебаний маятника, возникает характерное явление резонанса. Все эти физические выводы полностью соответствуют теории обыкновенных дифференциальных уравнений: $m\frac{d^2x}{dt^2} = -kx + F(t)$. В теории уравнений в частных производных оказывается, что тело имеет много частот и ведет себя как набор, как совокупность многих маятников с разными частотами, имеет много резонансов [2]. Если у Вас дома есть рояль или пианино — проверьте сказанное. Медленно, беззвучно опустите клавишу, чтобы только освободить струну, не ударяя ее молоточком. Теперь резко ударьте по другой клавише — и слушайте, как откликается свободная струна. . .

Для вещества или поля, заполняющего все пространство, характерен другой тип решения уравнений в частных производных — распространение волн. В частности, такими волнами в электромагнитной теории являются радио и свет, в упругих средах — звук. Замечательна способность волн переносить информацию: давление или электрическое поле в одной точке (вблизи приемника) как функция времени оказывается подобным зависимости от времени той же величины вблизи источника-передатчика.

Можно построить решения уравнений, описывающие направленный луч прожектора или лазера. Бросается в глаза сходство между лучом прожектора и струей из брандспойта. Знание свойств решений задач разных типов и аналогии между явлениями, описываемыми однотипными уравнениями, имели и имеют важнейшее значение для развития физики.

Метод *математического моделирования* состоит в отыскании математиче-

---

[2] Если Вы хотите понять сказанное глубже, просмотрите снова § 10.8.

ской схемы, настолько близкой к рассматриваемому явлению, что изучение явления можно заменить рассмотрением этой схемы. Такой математической схемой может быть обыкновенное дифференциальное уравнение или система таких уравнений. Иногда это — дифференциальное уравнение в частных производных с теми или иными начальными и граничными условиями (ср. § 10.8). В иных, более сложных математических конструкциях приходится иметь дело с понятием *вероятности*, о котором мы еще скажем ниже. Существенно, чтобы модель достаточно точно описывала данное явление реальной жизни [3]. Если мы придем при этом к уже изученной ранее математической схеме, то можем сразу же перенести все относящиеся к ней результаты на новый случай. Так, например, из того, что механические (см. гл. 10) и электрические колебания (гл. 13) описываются однотипными дифференциальными уравнениями, вытекает наличие в теории электрических цепей ряда характерных для механических колебаний фактов, например явления резонанса.

Сила математики в значительной степени связана с повторяемостью одних и тех же математических схем, приложимых к большому числу разнородных явлений: математика представляет собой, так сказать, достаточно ограниченный набор «ключей», каждый из которых неожиданным образом отпирает множество, казалось бы, вовсе не похожих друг на друга дверей (ср., например, § 2.1, 2.2 и 2.5). Это обстоятельство связано с тем, что природа устроена просто, без специальных ухищрений, — а простых математических конструкций существует не так уж много. Огромная общность математических методов давно уже вызывает благодарное изумление естествоиспытателей [4].

Спектры атомов долго составляли загадку для физиков: загадкой был прежде всего сам факт, что один и тот же атом испускает или резонансно поглощает колебания нескольких различных, но вполне определенных частот. Сходство с колебаниями упругих тел позволило подойти к формулировке уравнений *квантовой механики*. Сходство между потоком частиц и решениями уравнений, характерных для волновых явлений, также нашло свое применение в квантовой механике, породив своеобразный корпускулярно-волновой дуализм. Так, свет можно рассматривать как волну (точка зрения Х. Гюйгенса) и одновременно как поток частиц (или корпускул) света — фотонов (концепция И. Ньютона, ср. § 12.6 и след.). До начала нашего века казалось, что эти два подхода противоречат один другому и что нужно выбрать какой-то один из них. Современная квантовая механика показывает, как и в каком смысле верны (и дополняют одна другую) обе точки зрения.

Выше мы привели несколько примеров, относящихся к теории уравнений того или иного типа.

Иной пример доставляют нам комплексные числа и функции комплексной переменной (см. гл. 14, 15 и 17). Первоначально возникшие из задач решения алгебраических уравнений, комплексные числа долго вызывали большое недоверие. Однако прогресс гидро- и аэромеханики оказался тесно связанным с этими необычными «числами». При этом теория функций комплексной переменной была создана О. Коши и Б. Риманом и далее развита Карлом Вейерштрассом (1815—1897) как раз тогда, когда этот аппарат стал необходим для дальнейшего развития науки и техники. Создание воздухоплавания в начале нашего века существенно опиралось на расчет потоков жидкости или газа методами «комплексного анализа». Пионером этого направления явился замечательный русский механик Николай Егорович Жуковский (1847—1921).

---

[3] Слова «достаточно точно» в последней фразе подчеркивают неизбежность *идеализации* (т. е. огрубления рассматриваемого явления, отбрасывания второстепенных и в данный момент не интересующих нас деталей), с которой мы всегда встречаемся при переходе от физической реальности к описывающей ее математической схеме.

[4] См., например, статьи известных физиков, лауреатов Нобелевской премии: *Вигнер Е.* Непостижимая эффективность математики в естественных науках. — В кн.: Этюды о симметрии. М.: Мир, 1970, с. 182—197; *Янг Ч.* Эйнштейн и физика второй половины XX века. — Успехи физ. наук, 1980, т. 132, вып. 1, с. 169—175.

Другой замечательный пример развития математики дает геометрия.

Весь повседневный опыт учит нас, что в пространстве удобно ввести три координаты x, y, z. Кажется, что дальнейшие хитросплетения не нужны, они «от лукавого». Между тем можно ведь и иначе ввести координаты ξ, η, ζ, так, чтобы условие $\xi=\text{const}$ соответствовало какой-то кривой поверхности. Так, можно, например, ввести в пространстве три координаты ρ, φ и ϑ, где $\rho=OM$ — расстояние переменной точки M от начала координат O, а φ и ϑ — географические координаты (долгота и широта) на сфере $\rho=\text{const}$. (При этом поверхности $\varphi=\text{const}$ являются полуплоскостями, а поверхности $\vartheta=\text{const}$ — конусами (почему?).) Но можно вводить координаты и бесчисленными другими методами.

Итак, можно ввести *криволинейные координаты* ξ, η, ζ и затем с трудом и мучениями научиться вычислять расстояния между точками и другие величины с помощью новых координат.

На первый взгляд — это скучное, малополезное занятие. Нужно обладать очень специфическим складом ума, чтобы увидеть красоту в преодолении трудностей и развивать громоздкую теорию, использующую любые, самые общие координаты [5].

Разработка теории криволинейных координат в евклидовом пространстве кажется чисто методическим достижением. Однако при этом обобщенные координаты ξ, η, ζ одинаково удобны (или одинаково неудобны) как для описания обычного пространства, в котором верна геометрия Евклида, так и для описания «искривленного» пространства. Прямоугольные координаты x, y, z очень удобны для обычного пространства, но абсолютно непригодны для описания кривого пространства: эти координаты не дают даже намека на саму возможность существования каких-то других пространств.

Переход к криволинейным координатам, казавшийся ненужным осложнением, подготавливает нас к выходу в область изучения таких пространств, о самом существовании которых мы ранее не подозревали. Потом оказывается, что сила всемирного тяготения связана с тем, что пространство немного искривлено. Впрочем, здесь слово «немного» относится к условиям на Земле, в солнечной системе. В явлениях большого масштаба (катастрофических взрывах звезд, в эволюции Вселенной) искривленность пространства может ощущаться очень сильно — и здесь уже неизбежен переход от обычной евклидовой геометрии к неевклидовой. Поэтически можно сказать, что создание неевклидовой геометрии — чисто математическое достижение [6] — было той вспышкой молнии, за которой последовал громовой удар — создание общей теории относительности, т. е. геометрической теории тяготения. А в настоящее время в физику бурно вторгаются и совсем новые разделы геометрии, на первый взгляд кажущиеся еще более далекими от всякой реальности.

Еще в XVIII в. Д'Аламбер в статье «Размерность» (в прославленной «Энциклопедии» Дидро и Д'Аламбера) обращал внимание на то, что наш мир, по существу, является *четырехмерным*: наряду с тремя пространственными координатами x, y, z точки для различения происходящих в мире событий важно знать еще и время t. При этом связь между координатами x, y, z, t четырехмерного физического «мира событий» впоследствии оказалась более сложной, чем первоначально представляли себе Ньютон и Д'Аламбер, в частности такой, что в этом «мире» нельзя однозначно различать пространственные координаты x, y, z и временную координату t (в этом заключается сущ-

[5] Именно таким умом обладал великий Гаусс, разработавший «геометрию в криволинейных координатах»: с глубоким пониманием прикладного значения математики (так, к криволинейным координатам Гаусс пришел от своих занятий геодезией) он соединял неистребимое любопытство к тайнам математического мира, побуждавшее его отдавать многие часы и дни трудоемким вычислениям в надежде проникнуть в эти тайны (скажем, переводу простых дробей в десятичные с сотнями цифр в периоде дроби).

[6] Здесь снова можно удивляться тому, как уместно появились у математиков (в трудах Николая Ивановича Лобачевского из Казани (1792—1856), венгра Яноша Бойаи (1802—1860), К. Ф. Гаусса, несколько позднее — Б. Римана, еще позднее — немца Феликса Клейна (1849—1925) и француза Анри Пуанкаре (1854—1912)) необходимые для прогресса физики неевклидовы пространства!

ность *частной теории относительности Эйнштейна*). Позже в теоретическую физику (по существу, начиная с трудов Лагранжа, младшего современника Д'Аламбера) вошли *многомерные* пространства, например так называемое *фазовое пространство*: для простейшего случая движущейся в пространстве точки фазовое пространство будет шестимерным с координатами x, y, z, x', y', z' ($x'=\frac{dx}{dt}$, $y'=\frac{dy}{dt}$, $z'=\frac{dz}{dt}$ — это проекции скорости точки [7]). И сейчас в теоретической физике нам сплошь и рядом приходится иметь дело с многомерным миром, который и вообразить себе невозможно, или даже — страшно сказать — с «бесконечномерным» миром, каждая точка которого задается указанием бесконечного множества ее координат (чисел — и еще хорошо, если мы числа считаем вещественными, а не комплексными). В частности, с такими «бесконечномерными пространствами» мы встречаемся в квантовой механике — и весьма удачно, что теория бесконечномерных пространств была подготовлена знаменитым математиком Д. Гильбертом (1862—1943), к радости физиков, очень своевременно: совсем незадолго до создания квантовой механики.

В последние годы неожиданные приложения в физике нашла и идущая от А. Пуанкаре *топология* — один из новых и притом весьма абстрактных разделов геометрии. Топология имеет дело исключительно с самыми «грубыми» свойствами фигур. Так, например, шар и куб в топологии ничем не различаются, а баранка (тор) отлична от них, поскольку в баранке есть дыра. Однако эти «грубые» геометрические свойства топология анализирует весьма тщательно и глубоко — и этот нетривиальный анализ вдруг тоже заинтересовал физиков [8]. И никто не может сказать, какие еще новые подходы к пространству возьмут на вооружение физики будущих времен [9].

На скрещении (многомерного) математического анализа и топологии лежит недавно созданная (в первую очередь французским математиком Рене Томом (р. в 1923 г.)) *теория катастроф*, сразу получившая широкую (может быть, даже чересчур широкую) известность в связи с возможностями многообразных ее применений в естественных, гуманитарных и социальных науках [10]. В теории катастроф ставится вопрос о том, в каких условиях плавные зависимости и хорошие аналитические функции могут приводить к описанию явлений, подобных взрыву, т. е. к разрывной зависимости в ответе. Простейшие примеры такого рода были известны давно, еще до того, как появилось само название «теория катастроф».

Пусть величина a задана как гладкая функция b, $a=f(b)$, и пусть на практике можно задавать определенное значение a и наблюдать или измерять соответствующее b. Гладкость функции $f(b)$ отнюдь не гарантирует гладкости обратной функции $b=\varphi(a)$. В частности, если $f(b)$ имеет при $b=b_m$ максимум $a=a_{max}$, то при $a<a_{max}$ есть два значения b вблизи b_m, при $a>a_{max}$ — ни одного. При плавном изменении a величина b терпит разрыв.

С явлениями такого рода мы часто встречаемся в теории горения и взрывов, когда условия химической реакции определяются решениями алгебраических уравнений. Внешние условия

[7] Изображение физического процесса в фазовом пространстве рассматриваемой (механической) системы носит название «фазового портрета» процесса. Так, фазовый портрет равномерного движения $z=at+b$ точки на прямой — оси z — представляет собой прямую $z'=a$ в фазовой плоскости (z, z') движущейся точки $\left(z'=\frac{dz}{dt}\right.$ — скорость точки$\left.\right)$. Фазовый портрет гармонического колебания, характеризующегося постоянством (полной) энергии $E=\frac{kx^2}{2}+\frac{m(x')^2}{2}$, описывает точка в фазовой плоскости (x, x'), движущаяся по эллипсу $E=\text{const}$.

[8] Хочется указать читателю на превосходный (но отнюдь не легкий!) учебник: *Дубровин Б. А., Новиков С. П., Фоменко А. Т.* Современная геометрия (М.: Наука, 1979), в значительной степени ориентированный именно на физиков-теоретиков.

[9] Общий вопрос о взаимоотношении физики и математики с позиций сегодняшнего дня рассматривается в увлекательной, но не простой брошюре: *Манин Ю. И.* Математика и физика. М.: Знание, 1979. Этому же вопросу посвящены и названные в сноске 4 статьи Е. Вигнера и Ч. Янга.

[10] См., например, брошюру В. И. Арнольда «Теория катастроф» (М.: Знание, 1981) или обстоятельную (и богатую выразительными примерами) книгу Т. Постона и И. Стюарта «Теория катастроф и ее приложения» (М.: Мир, 1980).

входят как параметры, но решение уравнений может быть разрывной функцией параметров. Разрывы — это особенности решений. В теории катастроф выясняется, что эти особенности можно классифицировать, разбив их на несколько «грубых» (топологических) типов, отвечающих характерным явлениям тех жизненных феноменов, которые описываются рассматриваемым уравнением $u=f(z)$.

В исследованиях природы необходимо все: владение математическим аппаратом и преодоление математических трудностей, идейная смелость и физическая интуиция, умело поставленный эксперимент и математическое моделирование — все эти очень разные подходы равно необходимы, только сплав их ведет к продвижению вперед.

Абсолютно немыслима современная физика без понятия *вероятности*, принципиально отличного от всех понятий, рассматривавшихся в настоящей книге.

Все физические законы, о которых мы упоминали выше: второй закон Ньютона $mx''=F$ или его частный случай — закон инерции (см. § 9.4), закон Бойля—Мариотта $pV=\text{const}$ или его обобщение — закон ван дер Ваальса (см. § 7.1), имели так называемый *динамический* характер, т. е. позволяли точно предсказать то или иное явление: при уменьшении объема газа вдвое его давление увеличивается вдвое (закон Бойля—Мариотта); при движении тела в отсутствие действия сил скорость его в дальнейшем останется точно такой же, какой она была в начальный момент. Совсем иной характер имеют, скажем, законы падения подброшенной случайным образом монеты: здесь заранее можно только утверждать, что при многократном бросании монеты п р и м е р н о в п о л о в и н е с л у ч а е в она будет падать гербом кверху, т. е. что *вероятность* выпадения герба равна 0,5 (этот результат гарантируется условием «правильности» монеты, утверждающим полную равноправность обеих ее сторон). Физические законы, позволяющие предсказать не точный исход того или иного эксперимента, а лишь вероятность того или иного исхода, т. е. частоту, повторяемость этого исхода при многократном воспроизведении опыта в одних и тех же условиях, называют *статистическими* законами; раздел математики, занимающийся такими законами, носит название *теории вероятностей* [11].

Рассмотренный выше пример с многократным бросанием монеты представляет собой простейший почти тривиальный пример *вероятностного процесса*, характеризующегося случайным характером исходов его отдельных стадий (в данном случае — отдельных бросаний). Вот более типичный пример того же рода, использовавшийся физиками в качестве простой модели процесса диффузии газов (модель П. и Т. Эренфестов [12]).

Пусть мы имеем сосуд, разделенный на две части A и B пористой мембраной. В сосуде имеются N частиц (скажем, молекул газа), как-то распределенных по частям A и B; в каждый момент времени выбирается наугад одна из частиц N и перемещается из той части сосуда, в которой она находится, в другую часть. Можно ставить вопрос о том, за сколько актов в среднем начальное распределение — 100 частиц в A и 0 частиц в B — перейдет в состояние: 50 частиц в A и 50 в B. Оказывается, что надо около 140 актов. В принципе возможен и обратный процесс: при начальном состоянии 50 частиц в A и 50 в B случайные перемещения могут дать и 100 частиц в A, 0 в B, — но для этого потребуется $2^{100}\approx 10^{30}$ актов. Этот пример иллюстрирует понятия термодинамической необратимости и флюктуаций [13].

Но случай двух сосудов есть идеализация. Типичны задачи движения частиц в пространстве — поиски закона изменения распределения частиц в зависимости от координат и времени. Когда речь идет о крупных частицах, видимых в микроскоп, мы говорим о броунов-

---

[11] Простейшие понятия теории вероятностей излагаются в рассчитанных на начинающих книгах: *Гнеденко Б. В., Хинчин А. Я.* Элементарное введение в теорию вероятностей. М.: Наука, 1976; *Мостеллер Ф., Рурке Р., Томас Дж.* Вероятность. М.: Мир, 1969. Отметим также учебники: *Феллер В.* Введение в теорию вероятностей и ее приложения. М.: Мир, 1967. Ч. 1; *Нейман Ю.* Вводный курс теории вероятностей и математической статистики. М.: Наука, 1968; *Вентцель Е. С.* Теория вероятностей. М.: Наука, 1964.

[12] Пауль (Павел Сигизмундович) Э р е н ф е с т (1880—1933) и Татьяна Алексеевна А ф а н а с ь е в а - Э р е н ф е с т (1876—1964) — крупные физики XX века; работали в России (в СССР) и в Голландии.

[13] По поводу дальнейших деталей, а также доказательств сформулированных результатов см., например: *Кемени Дж., Снелл Дж.* Конечные цепи Маркова. М.: Наука, 1970, § 7.3.

ском движении. Характерно, что изучавшие броуновское движение или диффузию газов физики (А. Эйнштейн, М. Смолуховский, А. Д. Фоккер и М. Планк) обратились к учению о вероятностных процессах до того, как эта проблематика привлекла внимание математиков (А. Н. Колмогоров, А. Я. Хинчин, В. Феллер, П. Леви и др.). В связи с этим основное уравнение (непрерывных) процессов, родственных рассмотренному нами «игрушечному» процессу перекладывания шаров, называют и уравнением Эйнштейна—Смолуховского, и уравнением Фоккера—Планка, и уравнением Колмогорова.

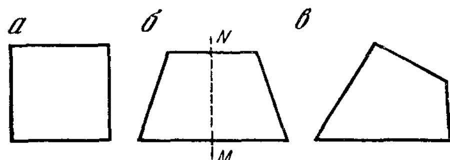


Рис. 3.1

Физике наших дней присуще обостренное внимание именно к статистическим законам природы, а следовательно, и широкое использование математического аппарата теории вероятностей. В частности, переход от классической механики Галилея и Ньютона к квантовой механике одним из своих аспектов имеет замену «динамического» мира рационалистов XVII—XIX вв. вероятностным, случайным миром современной физики микромира, где вообще принципиально нельзя говорить о точной траектории той или иной частицы, например электрона, но только — о вероятностях нахождения электрона в данный момент времени в той или иной точке пространства.

Наконец, самое первостепенное значение приобрел сегодня для современной физики еще один сравнительно молодой раздел математики, а именно *теория групп*. Эта теория изучает «степень симметрии» тех или иных (физических или других) объектов. Так, например, ясно, что квадрат (рис. 1, *а*) «симметричнее» равнобочной трапеции (рис. 1, *б*), а последняя — «более симметрична», чем произвольный четырехугольник (рис. 1, *в*). Точный смысл этих утверждений таков. Ясно, что *группа симметрии* квадрата, состоящая из всех его самосовмещений, более обширна, чем группа симметрии трапеции или произвольного четырехугольника. Группа симметрии квадрата содержит симметрии относительно всех его диагоналей и средних линий, а также повороты вокруг центра квадрата на 90, 180 и 270°, в то время как группа симметрии трапеции сводится к одной лишь симметрии относительно прямой MN (см. рис. 1, *б*), а группа симметрии четырехугольника (см. рис. 1, *в*) даже и еще беднее — она состоит из одного лишь «тождественного самосовмещения», оставляющего все точки на месте, и не содержит ни одного настоящего движения!

Создателем теории групп был гениальный юноша, француз-революционер Эварист Галуа (1811—1832), убитый на дуэли в возрасте 20 лет (эта дуэль была, видимо, подстроена полицией). Развивая в начале 30-х годов XIX в. идеи этой теории в применении к учению об алгебраических уравнениях (Галуа предлагал классифицировать уравнения в соответствии со свойственными им «группами симметрии»), он и не помышлял о приложении созданных им понятий к физике. Еще в начале нашего столетия известный английский астрофизик Дж. Джинс активно протестовал против включения элементов теории групп в курс математики для студентов-физиков, утверждая, что «уж эта-то теория никогда физикам не понадобится». В наши дни, когда теория групп приобрела в физике такое огромное значение, высказывание Джинса часто цитируют как пример полностью провалившегося предсказания.

Когда сегодня говорят о роли в физике теории групп, то в основном имеют в виду те применения, которые имеет эта теория в фундаментальных вопросах строения материи, в первую очередь — в теории элементарных частиц. Но первые приложения идей симметрии к практике возникли не в теории элементарных частиц, а в *кристаллографии* — ведь кристаллы представляют собой весьма симметричные тела. При этом оказалось, что кристаллы можно классифицировать с позиций присущей им симметрии и что эта классификация проливает существенный свет на все свойства кристаллов. Огромным событием явилось перечисление в конце XIX в. всех возможных систем симметрии кристаллов — их всего оказа-

лось 230. Это *федоровские* группы, называемые так по имени замечательного русского кристаллографа Евграфа Степановича Федорова (1853—1919), которому наряду с немцем А. Шенфлисом и англичанином В. Барлоу принадлежит открытие кристаллографических групп.

Физикам оказались нужны и так называемые *шубниковские* группы, носящие имя советского кристаллографа А. В. Шубникова (1887—1970), изучавшего симметрию цветных (например, черно-белых) орнаментов, которая учитывает и цвет отдельных участков: ведь ясно, что роль (черного или белого) цвета могут играть и другие физические характеристики рассматриваемых объектов, например, заряд.

Перечисление федоровских (или шубниковских) групп потребовало глубокого анализа самого понятия «системы симметрии», понимания смысла операции композиции поворотов, заменяющей несколько поворотов кристалла одним им равносильным.

Ясно, что каждое самосовмещение кристалла или иной фигуры — это определенное преобразование (движение) в пространстве; при этом композицию, т. е. последовательное выполнение двух преобразований α и β — сначала β, а затем α — мы можем назвать их произведением: $\gamma=\alpha\beta$. Так, композиция двух симметрий σ относительно прямой $0x$ или четырех поворотов π на 90° представляет собой «тождественное движение» ε, возвращающее каждую точку на место: $\sigma^2=\varepsilon$ и $\pi^4=\varepsilon$ (рис. 2). Таким путем мы приходим к «арифметике» (или «алгебре») симметрий, получаем возможность использовать в «исчислении симметрий» развитой аппарат математики; вот это-то «исчисление преобразований» (симметрий) и составляет предмет теории групп. Своеобразие ситуации здесь заключается в том, что в исчислении симметрий возможно неравенство $\alpha\beta \neq \beta\alpha$: рассматриваемые нами операции могут быть *некоммутативны*, т. е. *неперестановимы*[14]. Так, например, произведение $\sigma\pi$ поворота на 90° и симметрии относительно оси x (сначала поворот, а затем симметрия) равносильно одной симметрии относительно прямой $x+y=0$ (рис. 3, *а*), а произведение $\pi\sigma$ (сначала симметрия, а затем поворот) представляет уже иную операцию — симметрию относительно прямой $x—y=0$ (рис. 3, *б*).

Все эти соображения сыграли существенную роль в изучении симметрии кристаллов. Вспомните сказанное в § 14.1 о происхождении комплексных

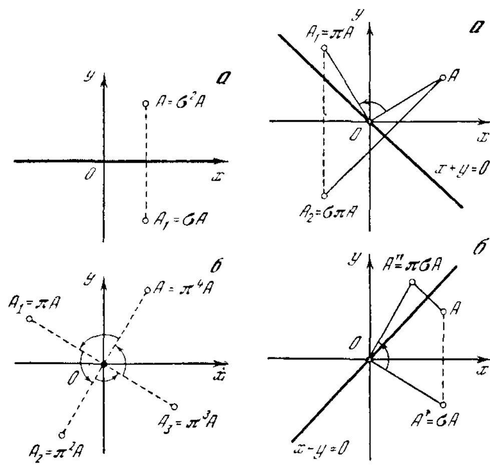


Рис. 3.2

Рис. 3.3

чисел! Человек начал считать с натуральных чисел 1, 2, 3. . . Впоследствии оказалось целесообразным рассмотреть также и отрицательные числа —1, —2, —3 и др., а также нуль; с другой стороны, понадобились дроби или даже иррациональные числа, такие, как $\sqrt{2}$. Дальнейшее обобщение понятия числа привело к комплексным числам; лишь после введения этих необычных чисел получила полное завер-

[14] Первой системой некоммутативных чисел явились так называемые *кватернионы*, *открытые знаменитым ирландским математиком* и механиком Уильямом Роуаном Гамильтоном (1805—1865). От комплексных чисел $x+iy=z$ кватернионы Гамильтона отличаются наличием трех комплексных единиц i, j, k (так что общий кватернион записывается как $w=u+xi+yj+zk$, где u, x, y, z — вещественные числа), квадрат каждой из которых равен —1, но которые не коммутируют между собой (так, $ij=k$, а $ji=-k$). Исчисление кватернионов Гамильтона явилось первым вариантом *векторного исчисления* (кватернион $w_0=u$ без комплексных единиц Гамильтон назвал *скаляром*, а чисто комплексный кватернион $w_1=xi+yj+zk$ с тремя компонентами x, y и z — *вектором*) — и прославленный «Трактат об электричестве и магнетизме» Дж. К. Максвелла (1873) был еще изложен не на привычном нам языке векторов, а на родственном ему языке кватернионов.

шение теория уравнений (в частности, элементарная теория квадратных уравнений). В гл. 14, 15, 17 было показано сколь плодотворным оказалось это обобщение понятия числа — особенно, когда удалось сочетать его с понятиями производной и интеграла.

Однако на коммутативность операций над числами ученые покушаться не решались — для всех вновь вводимых типов чисел всегда было $a+b=b+a$ и $ab=ba$! Поэтому шаг, который сделал Э. Галуа, надо счесть весьма принципиальным: ведь он впервые рассмотрел группы, элементы которых некоммутативны — в «арифметике Галуа» уже, вообще говоря, $ab \neq ba$. Во всем остальном теория кристаллографических групп близка к обычной арифметике: в ней существует единица, умножение на которую не меняет числа (в «исчислении симметрий» роль единицы играет «тождественная симметрия» ε); операция деления вводится как обратная умножению ($a : b=ab^{-1}=c$, если $a=cb$) и т. д. Однако то обстоятельство, что в группе умножение некоммутативно сыграло впоследствии для физики весьма важную роль.

В учении о кристаллах теория групп активно использовалась еще в XIX в.; однако от глубоких проблем, связанных с основными «строительными камнями мироздания» — элементарными частицами, возникающие здесь задачи казались весьма далекими. Но в 1932 г. знаменитый немецкий физик Вернер Гейзенберг установил, что замена положительно заряженного *протона* в ядре на аналогичную же электрически нейтральную частицу *нейтрон* математически подобна именно повороту. Это открытие явилось блестящей страницей истории физики! Собственно, с этого момента в теоретическую физику вошли группы и некоммутативные операции, вошел широко понимаемый принцип симметрии. И по мере того как физики открывали все новые и новые частицы, идея симметрии частиц и применения теории групп становились все более существенными. Постепенно используемая теория осложнялась: приходилось рассматривать повороты в воображаемом многомерном пространстве, учитывать заряд частиц и ее вращательный момент (спин). Подход к физическим объектам с позиции свойственной им симметрии сегодня является чуть ли не самым первым среди тех методов, с помощью которых физики пытаются разобраться во всем многообразии элементарных частиц, возникшем в результате многочисленных экспериментов и теоретических работ ученых второй половины XX в.

В современной физике широко используется не только теория *дискретных* (*кристаллографических*) *групп*, своеобразно копирующая элементарную арифметику (некоммутативная арифметика!), но и теория *непрерывных групп*, истоками которой служат алгебра и математический анализ. Наряду с группой самосовмещений квадрата (см. рис. 1, *а*) или кристалла можно также ставить вопрос о группе симметрии (группе самосовмещений) всей плоскости или (трех- или многомерного) пространства. Так, например, группа симметрии евклидовой плоскости состоит из всех ее движений:

$$x' = x\cos\varphi + y\cos\varphi + a,$$
$$y' = -x\sin\varphi + y\cos\varphi + b, \qquad (*)$$

где угол φ и числа a, b произвольны (ср. формулы (1.9.6)). То, что угол φ поворота и величины a и b переноса фигуры в плоскости могут быть любыми, позволяет говорить о б е с к о н е ч н о м а л о м повороте (на бесконечно малый угол $\Delta\varphi$ или $d\varphi$, где $d\varphi \to 0$) или переносе, что позволяет ввести понятие касательной к «групповой траектории» (определяемой, скажем, постоянством a и b и произвольным изменением φ), а также понятия дифференциала, производной и интеграла в групповом пространстве. Здесь можно также определить основные функции, вроде, например, экспоненты e^x (задаваемой, скажем, формулой (6.2.2), где x — «групповая переменная»). При этом некоммутативность величин, с которыми мы встречаемся в этой теории, создает совершенно новые проблемы, подобных которым нет ни в теории функций вещественной переменной (или нескольких вещественных переменных), ни в теории комплексных величин. Непрерывные группы, подобные группе (*) движений плоскости, впервые были подробно изучены норвежским математиком Софусом Мариусом Л и (1842—1899), по имени кото-

рого они сегодня называются *группами Ли*. В последние 10—20 лет теория групп Ли приобрела огромное значение для классификации элементарных частиц современной физики и для понимания взаимодействия этих частиц между собой.

Выдающееся значение теории групп и, в частности, групп Ли для современной физики побуждает нас кратко остановиться на истории происхождения этих разделов математики [15]. Об основополагающих, пионерских работах гениального Э. Галуа в области теории уравнений (предшественниками которого здесь явились знаменитый Ж. Л. Лагранж, итальянец Паоло Руффини (1765—1822) и замечательный норвежский математик Нильс Хенрик Абель (1802—1829) [16]) мы уже говорили выше. Работы Галуа не были своевременно оценены и даже просто замечены: не помогли и письма, написанные им знаменитым французским математикам, академикам О. Коши и Симеону Дени Пуассону (1781—1840). Пуассон оказался не в состоянии оценить идеи Галуа, значительно обогнавшие свое время, а Коши, видимо, даже и не прочел направленное ему безвестным юношей письмо.

Коши умер в 1857 г., и в 60-е годы прошлого века Французская академия наук приняла решение о публикации полного собрания его сочинений. Руководство этой работой было возложено на видного математика Камиля Мари Эдмона Жордана (1838—1922). В связи с этим поручением Жордан внимательно изучил все оставшиеся после Коши бумаги; среди них он обнаружил письмо Галуа, которое теперь (через 30 лет после написания!) произвело на Жордана огромное впечатление. Он тщательно изучил все принадлежащие Галуа работы, как опубликованные, так и оставшиеся в рукописях. Постепенно Жордан пришел к мысли о необходимости посвятить идеям Галуа специальную книгу, и эта монография: «Трактат о подстановках и алгебраических уравнениях», вышедшая в свет в 1870 г., явилась первым в мировой математической литературе учебником теории групп.

В период работы К. Жордана над «Трактатом о подстановках» у него учились двое друзей: молодые способные математики Ф. Клейн из Германии и С. Ли из Норвегии. Увлечение Жордана идеями Галуа и теорией групп полностью передалось его ученикам. Жордан впервые указал на существование двух различных типов групп: *дискретных* (*кристаллографических*) групп и *непрерывных* групп. Его ученики в известной мере поделили между собой эти две области математики. Основные достижения Ф. Клейна относятся к области учения о дискретных группах: выделенный им специальный тип таких групп (сегодня называемый *клейновыми группами*) вызывает в последнее время очень большой интерес. Наряду с этим Ф. Клейн в своей диссертационной работе, выполненной для вступления на профессорскую кафедру университета г. Эрлангена и впоследствии получившей наименование «Эрлангенской программы», указал на систему симметрии (группу симметрии) того или иного геометрического многообразия как на классификационный принцип, выделяющий отдельные ветви математической науки. Так, пространство Лобачевского, по Клейну, отличается от обычного пространства Евклида не только иными свойствами параллельных линий (второстепенный и мало что объясняющий признак!), но и совсем другой группой симметрии. Впоследствии этот принцип был распространен и на физику — и сегодня мы склонны считать, что, скажем, специальная теория относительности Эйнштейна отличается от классической механики Ньютона иным строением группы симметрии четырехмерного пространства—времени, причем в «мире Эйнштейна» эта группа устроена таким образом, что здесь ослаблено различие между пространственными и временно́й координатами (ср. § 9.8; см. книгу, названную в сноске 1 в § 9.8). Огромную роль играет этот клейновский принцип симметрии и для всей современной физики.

С. Ли представляет собой достаточно редкий в истории науки пример ученого-однолюба: вся его жизнь и вся колоссальная научная продукция (6 очень больших книг и множество статей, впоследствии собранных в многотомном «Собрании сочинений») была посвящена теории непрерывных групп (групп Ли). Учение о таких группах было разработано Ли с большой полнотой. В частности, Ли принадлежит перенесение на дифференциаль-

---

[15] Подробнее см., например: *Яглом И. М.* Феликс Клейн и Софус Ли. М.: Знание, 1977.

[16] Различие между работами Галуа и его предшественников заключалось не только (и не столько) в том, что Галуа впервые ввел сам термин «группа» и строго определил это понятие, но и в том, что Лагранж, Руффини и Абель работали исключительно с так называемыми *коммутативными*, или *абелевыми*, группами, для любых двух элементов a, b которых $ab=ba$, в то время как у Галуа появились группы общего вида (некоммутативные).

ные уравнения результатов Галуа, относящихся к алгебраическим уравнениям: основным различием здесь явилось то, что группа симметрии алгебраического уравнения (*группа Галуа*) является конечной, в то время как группа симметрии дифференциального уравнения (*группа Ли*) является уже непрерывной. Созданная Ли теория групп симметрии дифференциальных уравнений (в частности, вопрос о классификации дифференциальных уравнений по их группам симметрии) в последние десятилетия привлекает большое внимание математиков и физиков.

Скончавшийся еще в прошлом веке Ли, к сожалению, не увидел бурного расцвета созданной им теории; в частности, полностью неизвестными ему остались физические приложения теории непрерывных групп. Теоретико-групповой период развития теоретической физики начался с глубоких работ лауреатов Нобелевской премии немецкого физика М. Борна, его ученика В. Гейзенберга и англичанина П. А. М. Дирака; у истоков этого периода стоят один из крупнейших математиков XX в. Г. Вейль [17], ученик неоднократно упоминавшегося выше Д. Гильберта [18], и физик Э. Вигнер.

Теория симметрии и теория групп играют огромную роль в самой глубокой и самой трудной части физики — в теории элементарных частиц.

Долгое время теория развивалась в одном направлении; более сложные объекты физики «раскалывали», обнаруживая составляющие их более простые «части». Так, оказалось, что молекулы состоят из атомов, атомы — из ядер и электронов, ядра — из протонов и нейтронов. Сегодня мы добавили бы сюда утверждение о том, что протоны и нейтроны состоят из кварков. Развитие физики напоминало изучение младенцем деревянной куклы-матрешки: под каждым новым слоем обнаруживался следующий. При этом каждый новый слой оказывался проще предшествующего: все бесчисленные типы молекул сводились к комбинации из примерно 100 типов атомов, которым отвечают около 1000 разных ядер. Затем все это множество ядер атомов было сведено к двум всего типам элементарных частиц — протонам и нейтронам. Однако начиная примерно с середины нашего века в развитии теоретической физики обнаружилась и противоположная тенденция — не к упрощению, а к усложнению картины Вселенной. В космических лучах, а также с помощью ускорителей удалось обнаружить очень большое — превышающее 100 — число элементарных частиц. Все эти частицы столь же «элементарны», как протон или нейтрон. Правда, здесь еще один раз сработала модель куклы-матрешки: большое число частиц удалось описать как составные, складывающиеся из простейших, пока не наблюдавшихся частиц — кварков. Но в настоящее время физики занимаются не только и не столько вопросом о том, есть ли «еще более элементарные» частицы. Установление симметрии, т. е. сходства между различными частицами, приводит к важнейшим выводам об их свойствах. С симметрией самым тесным образом связано взаимодействие частиц между собой, возможность превращения одних частиц в другие и остальные их свойства. Кварки, электроны, нейтрино — может быть, для раскрытия их тайн надо еще глубже войти в вопросы симметрии этих частиц. И сегодня эта теория играет во всех современных исследованиях самую первостепенную роль. С этим и связан огромный интерес современных физиков ко всем аспектам теории симметрии и в первую очередь к теории (дискретных и непрерывных) групп.

Недавно в одной серьезной статье мы встретили следующее примечание: «Теорию групп ныне преподают в средней школе; поэтому лица, окончившие школу недавно, могут пропустить следующий параграф. Однако более пожилым физикам его необходимо проработать, чтобы иметь возможность беседовать со своими детьми и учить студентов». Мы еще не решаемся — пока не решаемся — включить элементы теории групп в свой учебник математики для начинающих физиков, но в дальнейшем, может быть, и придется это сделать.

Совершенно новое положение сложилось в математике и математической физике с появлением компьютеров — карманных вычислительных приборов, настольных считающих устройств и больших ЭВМ (электронных вычисли-

[17] См.: *Яглом И. М.* Герман Вейль. М.: Знание, 1967.

[18] См. обстоятельную книгу: *Рид К.* Гильберт. М.: Наука, 1977.

тельных машин) или даже целых комплексов ЭВМ, соединяющих несколько механизмов, отвечающих за отдельные этапы работы: работу таких комплексов можно несколько условно сравнить с высшей нервной деятельностью человека, где разные функции выполняют левое полушарие головного мозга (отвечающее, например, за речь) и правое (управляющее зрительными впечатлениями). Создание ЭВМ во многом изменило подход к задачам высшей математики, сделав иногда приближенный «лобовой» счет более простым, чем использование более сложных математических методов. Оно же породило возникновение совершенно новых разделов математики.

Иногда говорят, что математика — это мельница, которая перемалывает только то, что в нее заложено. Так оправдывают плохие результаты применения математики к неправильным исходным предположениям. В действительности эта мельница часто выдает больше того, что было заложено, выдает то, что не ожидалось!

В принципе математику можно рассматривать как разновидность уточненной, усовершенствованной логики. Замечательно, что, построив правила этой логики и выучив их, человек получил орудие гораздо более мощное, чем обыкновенный «здравый смысл», основаный на традиционной, «домашней» логике.

Человек руками создает простые орудия, применяя их, строит станки, с помощью которых создает еще более совершенные и сложные механизмы — и с помощью этих механизмов он способен сделать то, что недоступно голым рукам. Вот так же точно и математика, развивая все более сложные теории и создавая новые понятия, дает возможность овладеть самыми необычными явлениями природы.

В конце курса математической физики можно будет снова написать главу под названием «А что еще дальше?» — но не надо отчаиваться, потому что здесь уже близок тот рубеж, за которым продвижение вперед представляет собой не только учебу, но и творчество, развитие новых теорий.

Мы пытаемся представить себе читателя, который через несколько минут со вздохом облегчения закроет эту книгу. Наверно, большинство из Вас — школьники-старшеклассники или студенты-первокурсники. Пусть математика для Вас навсегда останется точным и прекрасным языком, способом выражения мыслей и способом мышления; пусть математика не будет предметом, который нужно на экзамене «весь сдать и ничего себе не оставить». Любите математику — и любовь будет взаимной, математика всегда поможет Вам в Вашей работе.

ЛИТЕРАТУРА

Пособия по высшей математике для начинающих

1. *Колмогоров А. Н., Абрамов А. М., Вейц Б. Е.* и др. Алгебра и начала анализа: Учебное пособие для 9—10 кл. средней школы. М.: Просвещение, 1980. 335 с.
2. *Понтрягин Л. С.* Математический анализ для школьников. М.: Наука, 1980. 86 с.
3. *Никольский С. М.* Элементы математического анализа. М.: Наука, 1981. 159 с.
4. *Лоренц Г. А.* Элементы высшей математики. М.; Л.: ГИЗ, 1926. Т. 1. 558 с.; Т. 2. 528 с. (на титульном листе последнего издания книги указаны авторы: Г. Лоренц, В. Шереметевский).
5. *Фихтенгольц Г. М.* Математика для техников. М; Л.: ГИЗ, 1926. 343 с.
6. *Выгодский М. Я.* Высшая математика для техникумов. М.: Высшая школа, 1969. 446 с.
7. *Выгодский М. Я.* Справочник по высшей математике. М.: Наука, 1977. 871 с.
8. *Курант Р., Роббинс Г.* Что такое математика? М.: Просвещение, 1967, гл. VIII (с. 431—525).

Более сложные учебники по высшей математике

9. *Мышкис А. Д.* Лекции по высшей математике. М.: Наука, 1973. 640 с.
10. *Натансон И. П.* Краткий курс высшей математики. М.: Наука, 1968. 727 с.
11. *Смирнов В. И.* Курс высшей математики. М.: Наука, 1974. Т. 1. 474 с.
12. *Берс Л.* Математический анализ. М.: Высшая школа, 1975. Т. 1. 519 с; Т. 2. 544 с.
13. *Фихтенгольц Г. М.* Математика для инженеров. Л.; М.: Гостехиздат, 1933, Т. 1. 488 с.; Т. 2. 330 с.
14. *Выгодский М. Я.* Основы исчисления бесконечно малых. М.; Л.: Гостехиздат, 1933. 464 с.

Дополнительная литература по математике

15. *Зельдович Я. Б., Мышкис А. Д.* Элементы прикладной математики. М.: Наука, 1972. 592 с.
16. *Курант Р.* Курс дифференциального и интегрального исчисления. М.: Наука, 1967. Т. 1. 704 с.; 1970, Т. 2. 671 с.
17. *Мышкис А. Д.* Математика для ВТУЗов: Специальные курсы. М.: Наука, 1971. 632 с.
18. *Смирнов В. И.* Курс высшей математики. М.: Наука, 1974. Т. 2. 655 с.; Т. 3. Ч. 1. 323 с.; Ч. 2. 672 с.; 1958. Т. 4. 812 с.; 1959, Т. 5. 655 с.
19. Математика в современном мире. М.: Мир, 1967. 205 с.
20. *Стюарт Я.* Концепция современной математики. Минск: Вышэйшая школа, 1980. 382 с.
21. *Курант Р., Роббинс Г.* Что такое математика? М.: Просвещение, 1967, гл. I—VII (с. 19—430).

Пособия по физике для начинающих

22. Элементарный учебник физики/ Под ред. акад. Г. С. Ландсберга. М.: Наука, 1975. Т. 1. 654 с.; Т. 2. 526 с.; Т. 3. 638 с.
23. *Марион Дж. Б.* Физика и физический мир. М.: Мир, 1975. 623 с.
24. *Ландау Л. Д., Китайгородский А. И.* Физика для всех. М.: Наука, 1978. Т. 1. 205 с.; Т. 2. 207 с.; 1979. Т. 3. 207 с.; Т. 4. 207 с.
25. *Орир Дж.* Популярная физика. М.: Мир, 1969. 556 с.
26. *Эллиот Л., Уилкокс У.* Физика. М.: Наука, 1975. 736 с.
27. *Кемпфер Ф.* Путь в современную физику. М.: Мир, 1972. 375 с.
28. *Линднер Г.* Картины современной физики. М.: Мир, 1977. 272 с.
29. *Пайерлс Р.* Законы природы. М.: Физматгиз, 1962. 340 с.
30. *Купер Л. Н.* Физика для всех. М.: Мир, 1973. Т. 1. 479 с.; 1974. Т. 2. 383 с.
31. *Роджерс Э. М.* Физика для любознательных М.: Мир, 1972. Т. 1. 468 с.; Т. 2. 652 с.; 1973, Т. 3. 644 с.
32. *Акоста В., Кован К., Грэм Б.* Основы современной физики. М.: Просвещение, 1981. 495 с.
33. *Поль Р. В.* Механика, акустика и учение о теплоте. М.: Наука, 1971. 479 с.; Учение об электричестве. М.: Физматгиз, 1962. 516 с.

Более сложная литература по физике [1]

34. Феймановские лекции по физике. М.: Мир, 1977, вып. 1—2. 438 с.; вып. 3—4. 495 с.; вып. 5. 299 с.; вып. 6. 346 с.; вып. 7. 287 с.; 1978, вып. 8. 523 с.; вып. 9. 541 с.
35. *Фейман Р.* Характер физических законов. М.: Мир, 1968. 232 с.
36. Берклеевский курс физики. В 5-ти томах. М.: Наука. Т. 1. *Киттель Ч., Найт У., Рудерман М.* Механика. 1975. 479 с.; Т. 2. *Парселл Э.* Электричество и магнетизм. 1975. 439 с.; Т. 3. *Крауфорд Ф.* Волны. 1976. 527 с.; Т. 4. *Вихман Э.* Квантовая физика. 1977. 415 с.; Т. 5. *Рейф Ф.* Статистическая физика. 1977. 351 с.
37. *Копылов Г. И.* Всего лишь кинематика. М.: Наука, 1981. 175 с.

Литература по истории науки

38. *Стройк Д. Я.* Краткий очерк истории математики. М.: Наука, 1978. 335 с.
39. *Гиндикин С. Г.* Рассказы о физиках и математиках. М.: Наука, 1981. 191 с.
40. *Белл Э. Т.* Творцы математики. М.: Просвещение, 1979. 255 с.

---

[1] «Продолжением» книг [34, 36] может служить знаменитый многотомный курс «Теоретической физики» академиков Л. Д. Ландау и Е. М. Лифшица.

ПРИЛОЖЕНИЯ

I. Таблица производных

1. $y = c, \quad \frac{dy}{dx} = 0.$

2. $y = x, \quad \frac{dy}{dx} = 1.$

3. $y = x^a, \quad \frac{dy}{dx} = ax^{a-1} = \frac{ay}{x}.$

4. $y = e^x, \quad \frac{dy}{dx} = e^x.$

5. $y = a^x, \quad \frac{dy}{dx} = a^x \ln a \approx 2{,}3a^x \lg a.$

6. $y = \ln x, \quad \frac{dy}{dx} = \frac{1}{x}.$

7. $y = \log_a x, \quad \frac{dy}{dx} = \frac{1}{x \ln a} \approx \frac{0{,}434}{\lg a} \frac{1}{x}.$

8. $y = \sin x, \quad \frac{dy}{dx} = \cos x.$

9. $y = \cos x, \quad \frac{dy}{dx} = -\sin x.$

10. $y = \operatorname{tg} x, \quad \frac{dy}{dx} = \frac{1}{\cos^2 x}.$

11. $y = \operatorname{ctg} x, \quad \frac{dy}{dx} = -\frac{1}{\sin^2 x}.$

12. $y = \arcsin x. \quad \frac{dy}{dx} = \frac{1}{\sqrt{1 - x^2}}.$

13. $y = \arccos x, \quad \frac{dy}{dx} = -\frac{1}{\sqrt{1 - x^2}}.$

14. $y = \operatorname{arc\,tg} x, \quad \frac{dy}{dx} = \frac{1}{1 + x^2}.$

15. $y = \operatorname{arc\,ctg} x, \quad \frac{dy}{dx} = -\frac{1}{1 + x^2}.$

II. Интегралы от некоторых функций

1. $\int dx = x + C.$

2. $\int x^a dx = \frac{x^{a+1}}{a+1} + C \quad (a \neq -1).$

3. $\int \frac{dx}{x} = \ln |x| + C.$

4. $\int \frac{dx}{ax + b} = \frac{1}{a} \ln |ax + b| + C.$

5. $\int a^x dx = \frac{a^x}{\ln a} + C.$

6. $\int e^{kx} dx = \frac{1}{k} e^{kx} + C.$

7. $\int x^n e^{kx} dx = \frac{1}{k} x^n e^{kx} - \frac{n}{k} \int x^{n-1} e^{kx} dx.$

8. $\int \frac{dx}{1 + e^{kx}} = \frac{1}{k} \ln \frac{e^{kx}}{1 + e^{kx}} + C.$

9. $\int e^{kx} \sin ax dx =$
$= \frac{e^{kx}}{k^2 + a^2} (k \sin ax - a \cos ax) + C.$

10. $\int e^{kx} \cos ax dx =$
$= \frac{e^{kx}}{k^2 + a^2} (k \cos ax + a \sin ax) + C.$

11. $\int \sin kx dx = -\frac{1}{k} \cos kx + C.$

12. $\int \cos kx dx = \frac{1}{k} \sin kx + C.$

13. $\int \frac{dx}{\sin^2 kx} = -\frac{1}{k} \operatorname{ctg} kx + C.$

14. $\int \frac{dx}{\cos^2 kx} = \frac{1}{k} \operatorname{tg} kx + C.$

15. $\int \sin^2 kx dx =$
$= \frac{x}{2} - \frac{1}{4k} \sin 2kx + C.$

16. $\int \cos^2 kx dx = \frac{x}{2} + \frac{1}{4k} \sin 2kx + C.$

17. $\int x^n \sin kx dx =$
$= -\frac{x^n}{k} \cos kx + \frac{n}{k} \int x^{n-1} \cos kx dx.$

18. $\int x^n \cos kx dx =$
$= \frac{x^n}{k} \sin kx - \frac{n}{k} \int x^{n-1} \sin kx dx.$

19. $\int \sin kx \sin lx dx =$
$= \frac{\sin (k - l) x}{2 (k - l)} - \frac{\sin (k + l) x}{2 (k + l)} + C,$
если $|k| \neq |l|$ (если $|k| = |l|$, см. № 15).

20. $\int \cos kx \cos lx dx =$
$= \frac{\sin (k - l) x}{2 (k - l)} + \frac{\sin (k + l) x}{2 (k + l)} + C,$
если $|k| \neq |l|$ (если $|k| = |l|$, см. №16).

21. $\int \sin kx \cos lx dx =$
$= -\frac{\cos (k + l) x}{2 (k + l)} - \frac{\cos (k - l)}{2 (k - l)} + C,$
если $|k| \neq |l|$.

22. $\int \operatorname{tg} kx\,dx = -\frac{1}{k}\ln|\cos kx| + C.$

23. $\int \operatorname{ctg} kx\,dx = \frac{1}{k}\ln|\sin kx| + C.$

24. $\int \sqrt{ax+b}\,dx = \frac{2}{3a}(ax+b)^{3/2} + C.$

25. $\int \frac{dx}{\sqrt{ax+b}} = \frac{2\sqrt{ax+b}}{a} + C.$

26. $\int \frac{dx}{\sqrt{a^2-x^2}} = \arcsin\frac{x}{a} + C.$

27. $\int x\sqrt{ax+b}\,dx =$

$= \frac{2(3ax-2b)(ax+b)^{3/2}}{15a^2} + C.$

28. $\int \sqrt{a^2-x^2}\,dx =$

$= \frac{1}{2}\left[x\sqrt{a^2-x^2} + a^2\arcsin\frac{x}{a}\right] + C.$

29. $\int \frac{\sqrt{a^2-x^2}}{x}\,dx =$

$= \sqrt{a^2-x^2} - a\ln\left|\frac{a+\sqrt{a^2-x^2}}{x}\right| + C.$

30. $\int x\sqrt{x^2+m}\,dx = \frac{1}{3}(x^2+m)^{3/2} + C.$

31. $\int \frac{dx}{\sqrt{x^2+m}} = \ln\left|x+\sqrt{x^2+m}\right| + C.$

32. $\int \frac{\sqrt{a^2+x^2}}{x}\,dx =$

$= \sqrt{a^2+x^2} - a\ln\left|\frac{a+\sqrt{a^2+x^2}}{x}\right| + C.$

33. $\int \sqrt{x^2+m}\,dx = \frac{1}{2}\left[x\sqrt{x^2+m} +\right.$

$\left.+ m\ln\left|x+\sqrt{x^2+m}\right|\right] + C.$

34. $\int \frac{\sqrt{x^2-a^2}}{x}\,dx =$

$= \sqrt{x^2-a^2} - a\arccos\frac{a}{x} + C.$

35. $\int \frac{dx}{x^2-a^2} = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right| + C.$

36. $\int \frac{dx}{x^2+a^2} = \frac{1}{a}\operatorname{arctg}\frac{x}{a} + C.$

37. $\int \frac{dx}{ax^2+bx+c} =$

$= \frac{2}{\sqrt{4ac-b^2}}\operatorname{arctg}\frac{2ax+b}{\sqrt{4ac-b^2}} + C,$

если $4ac-b^2>0$;

$\int \frac{dx}{ax^2+bx+c} = \frac{1}{\sqrt{b^2-4ac}} \times$

$\times \ln\left|\frac{2ax+b-\sqrt{b^2-4ac}}{2ax+b+\sqrt{b^2-4ac}}\right| + C,$

если $4ac-b^2<0$;

$\int \frac{dx}{ax^2+bx+c} = -\frac{2}{2ax+b} + C,$

если $4ac-b^2=0$, т. е. $c=\frac{b^2}{4a}$.

38. $\int \frac{dx}{(ax^2+bx+c)^n} =$

$= \frac{2ax+b}{(n-1)(4ac-b^2)(ax^2+bx+c)^{n-1}} +$

$+ \frac{(2n-3)\cdot 2a}{(n-1)(4ac-b^2)}\int \frac{dx}{(ax^2+bx+c)^{n-1}}$

при $n \geqslant 2$ и $4ac-b^2 \neq 0$.

39. $\int \frac{x\,dx}{ax^2+bx+c} = \frac{1}{2a}\ln|ax^2+bx+c| -$

$-\frac{b}{2a}\int \frac{dx}{ax^2+bx+c}$ (см. № 37).

40. $\int \frac{dx}{x(ax^2+bx+c)} =$

$= \frac{1}{2c}\ln\frac{x^2}{|ax^2+bx+c|} -$

$-\frac{b}{2c}\int \frac{dx}{ax^2+bx+c}$ (см. № 37).

41. $\int \frac{dx}{x^m(ax^2+bx+c)^n} =$

$= -\frac{1}{(m-1)cx^{m-1}(ax^2+bx+c)^{n-1}} -$

$-\frac{(2n+m-3)a}{(m-1)c}\int \frac{dx}{x^{m-2}(ax^2+bx+c)^n} -$

$-\frac{(n+m-2)b}{(m-1)c}\int \frac{dx}{x^{m-1}(ax^2+bx+c)^n}$

(при $m>1$).

42. $\int \sqrt[n]{ax+b}\,dx =$

$= \frac{n(ax+b)}{a(n+1)}\sqrt[n]{ax+b} + C.$

43. $\int \frac{dx}{\sqrt[n]{ax+b}} =$

$= \frac{n(ax+b)}{(n-1)a}\,\frac{1}{\sqrt[n]{ax+b}} + C.$

44. $\int \ln x\,dx = x\ln x - x + C.$

45. $\int (\ln x)^n\,dx = x(\ln x)^n - n\int(\ln x)^{n-1}dx.$

46. $\int \arcsin\frac{x}{a}\,dx =$

$= x\arcsin\frac{x}{a} + \sqrt{a^2-x^2} + C.$

47. $\int \arccos\frac{x}{a}\,dx =$

$= x\arccos\frac{x}{a} - \sqrt{a^2-x^2} + C.$

48. $\int \operatorname{arc\,tg} \frac{x}{a}\,dx =$
$= x \operatorname{arc\,tg} \frac{x}{a} - \frac{a}{2} \ln (a^2 + x^2) + C.$

49. $\int \operatorname{arc\,ctg} \frac{x}{a}\,dx =$
$= x \operatorname{arc\,ctg} \frac{x}{a} + \frac{a}{2} \ln (a^2 + x^2) + C.$

III. Некоторые разложения функций в ряды

1. $(1+x)^m = 1 + mx + \frac{m(m-1)}{2!} x^2 + \frac{m(m-1)(m-2)}{3!} x^3 + \ldots$
$(-1 < x < 1$, если m — не натуральное)

2. $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$
(x — любое).

3. $\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \ldots$
(x — любое).

4. $\operatorname{tg} x = x + \frac{1}{3} x^3 + \frac{2}{15} x^5 + \frac{17}{315} x^7 + \frac{62}{2835} x^9 + \ldots \quad \left(-\frac{\pi}{2} < x < \frac{\pi}{2}\right).$

5. $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots$
(x — любое).

6. $\ln (1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$
$(-1 < x \leqslant 1).$

7. $\arcsin x =$
$= x + \frac{x^3}{2 \cdot 3} + \frac{1 \cdot 3}{2 \cdot 4 \cdot 5} x^5 + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6 \cdot 7} x^7 + \ldots \quad (-1 \leqslant x \leqslant 1).$

8. $\operatorname{arc\,tg} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \ldots$
$(-1 < x < 1).$

IV. Некоторые числовые таблицы

Таблица 1

x	e^x	e^{-x}	x	e^x	e^{-x}	x	e^x	e^{-x}
0	1,000	1,000	1,5	4,482	0,223	3,8	44,701	0,0224
0,1	1,105	0,905	1,6	4,953	0,202	4,0	54,598	0,0183
0,2	1,221	0,819	1,7	5,474	0,183	4,5	90,017	0,0111
0,3	1,350	0,741	1,8	6,050	0,165	5,0	148,41	0,00674
0,4	0,492	0,670	1,9	6,686	0,150	5,5	244,69	0,00409
0,5	1,649	0,607	2,0	7,389	0,135	6,0	403,43	0,00248
0,6	1,822	0,549	2,2	9,025	0,1108	6,5	665,14	0,00150
0,7	2,014	0,497	2,4	11,023	0,0907	7,0	1 096,6	0,000912
0,8	2,226	0,449	2,6	13,464	0,0743	7,5	1 808,0	0,000553
0,9	2,460	0,407	2,8	16,445	0,0608	8,0	2 981,0	0,000335
1,0	2,718	0,368	3,0	20,086	0,0498	8,5	4 914,8	0,000203
1,1	3,004	0,333	3,2	24,533	0,0408	9,0	8 103,1	0,000123
1,2	3,320	0,301	3,4	29,964	0,0334	9,5	13 360	0,000075
1,3	3,669	0,273	3,6	36,598	0,0273	10,0	22 026	0,000045
1,4	4,055	0,247						

Таблица 2

x	$\ln x$	x	$\ln x$	x	$\ln x$	x	$\ln x$	x	$\ln x$	x	$\ln x$
1,0	0	1,5	0,405	2,2	0,788	3,2	1,163	5,0	1,609	7,5	2,015
1,1	0,0953	1,6	0,470	2,4	0,875	3,4	1,224	5,5	1,705	8,0	2,079
1,2	0,182	1,7	0,531	2,6	0,956	3,6	1,281	6,0	1,792	8,5	2,140
1,3	0,262	1,8	0,588	2,8	1,030	3,8	1,335	6,5	1,872	9,0	2,197
1,4	0,336	2,0	0,693	3,0	1,099	4,5	1,504	7,0	1,946	10,0	2,303

Таблица 3

x	$\sin x$	$\cos x$	$\operatorname{tg} x$	x	$\sin x$	$\cos x$	$\operatorname{tg} x$
0	0,000	1,000	0,000	3,2	—0,0584	—0,998	0,0585
0,1	0,0998	0,995	0,100	3,3	—0,158	—0,987	0,160
0,2	0,199	0,980	0,203	3,4	—0,256	—0,967	0,264
0,3	0,296	0,955	0,309	3,5	—0,361	—0,936	0,375
0,4	0,389	0,921	0,423	3,6	—0,443	—0,897	0,493
0,5	0,479	0,878	0,546	3,7	—0,530	—0,848	0,625
0,6	0,565	0,825	0,684	3,8	—0,612	—0,791	0,774
0,7	0,644	0,765	0,842	3,9	—0,688	—0,726	0,947
0,8	0,717	0,697	1,030	4,0	—0,757	—0,654	1,158
0,9	0,783	0,622	1,260	4,1	—0,818	—0,575	1,424
1,0	0,841	0,540	1,557	4,2	—0,872	—0,490	1,778
1,1	0,891	0,454	1,965	4,3	—0,916	—0,401	2,286
1,2	0,932	0,362	2,572	4,4	—0,952	—0,307	3,096
1,3	0,964	0,268	3,602	4,5	—0,978	—0,211	4,637
1,4	0,985	0,170	5,798	4,6	—0,994	—0,112	8,860
1,5	0,997	0,0707	14,101	4,7	—1,000	—0,0124	80,713
1,6	0,9996	—0,0292	—34,233	4,8	—0,996	0,0875	—11,385
1,7	0,992	—0,129	—7,697	4,9	—0,982	0,187	—5,267
1,8	0,974	—0,227	—4,286	5,0	—0,969	0,284	—3,881
1,9	0,946	—0,323	—2,927	5,1	—0,926	0,378	—2,449
2,0	0,909	—0,416	—2,185	5,2	—0,883	0,469	—1,886
2,1	0,863	—0,505	—1,710	5,3	—0,832	0,554	—1,501
2,2	0,808	—0,589	—1,374	5,4	—0,773	0,635	—1,218
2,3	0,746	—0,666	—1,119	5,5	—0,706	0,709	—0,996
2,4	0,675	—0,737	—0,916	5,6	—0,631	0,776	—0,814
2,5	0,598	—0,801	—0,747	5,7	—0,551	0,835	—0,660
2,6	0,516	—0,857	—0,602	5,8	—0,465	0,886	—0,525
2,7	0,427	—0,904	—0,473	5,9	—0,374	0,927	—0,403
2,8	0,335	—0,942	—0,356	6,0	—0,279	0,960	—0,291
2,9	0,239	—0,971	—0,246	6,1	—0,182	0,983	—0,185
3,0	0,141	—0,990	—0,143	6,2	—0,0831	0,997	—0,0834
3,1	0,0416	—0,999	—0,0416	6,3	—0,0168	1,000	0,0168

V. Международная система физических единиц СИ

Величина		Единица СИ	
наименование	размерность	наименование	обозначение
	Основные единицы		
Длина	м	метр	м
Масса	кг	килограмм	кг
Время	с	секунда	с
Сила электрического тока	А	ампер	А
Температура	К	градус Кельвина (кельвин)	К
Количество вещества	моль	моль	моль
Сила света	кд	кандела	кд
	Дополнительные единицы		
Плоский угол	рад	радиан	рад
Телесный угол	ср	стерадиан	ср
	Производные единицы		
Скорость	м/с	метр в секунду	м/с
Ускорение	м/с^2	метр на секунду в квадрате	м/с^2
Угловая скорость	рад/с	радиан в секунду	рад/с
Угловое ускорение	рад/с^2	радиан на секунду в квадрате	рад/с^2
Сила	кг·м/с^2	ньютон	Н
Давление	Н/м^2(=кг/(м·с^2))	паскаль	Па
Работа, энергия, количество теплоты	Н·м(=кг·м^2/с^2)	джоуль	Дж

Окончание

Величина		Единица СИ	
наименование	размерность	наименование	обозначение
Мощность	Дж/с(=кг·м²/с³)	ватт	Вт
Теплоемкость удельная	Дж/(кг·К)(=м²/(с²·К))	джоуль на килограмм-кельвин	Дж/(кг··К)
Импульс (количество движения)	кг·м/с	килограмм-метр в секунду	кг·м/с
Импульс силы	Н·с(=кг·м/с)	ньютон-секунда	Н·с
Частота	1/с(=с⁻¹)	герц	Гц
Количество электричества, электрический заряд	А·с	кулон	Кл
Электрическое напряжение, разность потенциалов, ЭДС	Вт/А(=кг·м²/(А·с³))	вольт	В
Напряженность электрического поля	В/м(=кг·м/(А·с³))	вольт на метр	В/м
Электрическое сопротивление	В/А(=Вт/А²=кг·м²/(А²·с³))	ом	Ом
Электрическая емкость	Кл/В(=А²·с⁴/(кг·м²))	фарада	Ф
Индуктивность	В·с/А(=кг·м²/(А²·с²)=Вб/А)	генри	Гн
Магнитный поток, магнитная индукция	$В\cdot с\left(=\frac{кг\cdot м^2}{А\cdot с^2}\right)$	вебер	Вб
Напряженность магнитного поля	А/м	ампер на метр	А/м
Плотность магнитного потока, магнитной индукции	Кл/м(=Вб/м²=А·с/м)	тесла	Тл
Магнитодвижущая сила	А	ампер	А

VI. Латинский и греческий алфавиты

Буква	Название	Буква	Название	Буква	Название	Буква	Название
			Латинский алфавит				
A a	а	*H h*	аш	*N n*	эн	*U u*	у
B b	бе	*I i*	и	*O o*	о	*V v*	ве
C c	це	*J j*	йот	*P p*	пэ	*W w*	дубль-ве
D d	де	*K k*	ка	*Q q*	ку	*X x*	икс
E e	е	*L l*	эль	*R r*	эр	*Y y*	игрек
F f	эф	*M m*	эм	*S s*	эс	*Z z*	зет
G g	ге			*T t*	тэ		
			Греческий алфавит				
Α α	альфа	Η η	эта	Ν ν	ню	Τ τ	тау
Β β	бета	Θ θϑ	тэта	Ξ ξ	кси	Υ υ	ипсилон
Γ γ	гамма	Ι ι	йота	Ο ο	омикрон	Φ φ	фи
Δ δ	дельта	Κ κ	каппа	Π π	пи	Χ χ	хи
Ε ε	эпсилон	Λ λ	ламбда	Ρ ρ	ро	Ψ ψ	пси
Ζ ζ	дзета	Μ μ	мю	Σ σ	сигма	Ω ω	омега

ОТВЕТЫ, УКАЗАНИЯ, РЕШЕНИЯ

Часть первая

Глава 1

1.2.4. $r=\sqrt{2}$, $\alpha=45°$; $r=2\sqrt{2}$, $\alpha=-45°$; $r=3\sqrt{2}$, $\alpha=-135°$; $r=4\sqrt{2}$, $\alpha=135°$. **1.2.5.** 2; $2\sqrt{2}$; $2\sqrt{2}$; $2\sqrt{2}$. **1.2.6.** $(0,\ a/\sqrt{2})$; $(a/\sqrt{2},\ 0)$; $(0,\ -a/\sqrt{2})$; $(-a/\sqrt{2},\ 0)$ (сделайте чертеж). **1.2.7.** $(a,\ 0)$; $(a/2,\ a\sqrt{3}/2)$; $(-a/2,\ a\sqrt{3}/2)$; $(-a,\ 0)$; $(-a/2,\ -a\sqrt{3}/2)$; $(a/2,\ -a\sqrt{3}/2)$ (сделайте чертеж). **1.2.8.** а) Два случая: $(-a/2,\ 0)$, $(a/2,\ 0)$, $(0,\ \pm a\sqrt{3}/2)$. б) Четыре случая: $(0,\ 0)$, $(a,\ 0)$, $(a/2,\ \pm a\sqrt{3}/2)$ и $(0,\ 0)$ $(-a,\ 0)$ $(-a/2,\ \pm a\sqrt{3}/2)$ (сделайте чертеж). **1.2.9.** $A_2(x_1,\ -y_1)$, $A_3(-x_1,\ y_1)$, $A_4(-x_1,\ -y_1)$ (сделайте чертеж).

1.3.3. а) $y=-x$; б) $y=2x-1$. **1.3.4.** Если прямая пересекает оси координат в точках M и N, то a и b, взятые с подходящими знаками, — длины отрезков OM и ON.

1.4.2. Воспользуйтесь тем, что $x^2-2x+2=(x-1)^2+1$; $2x^2+4x=2(x+1)^2-2$. **1.4.3.** а) Растяжение от начала координат (гомотетия) с коэффициентом $\sqrt{k}$ переводит точку $M(x,\ y)$ в точку $M'(x',\ y')$, где $x'=\sqrt{k}x$, $y'=\sqrt{k}y$. Отсюда $x'y'=kxy=k$. б) Растяжение в направлении оси x с коэффициентом k переводит точку $M(x,\ y)$

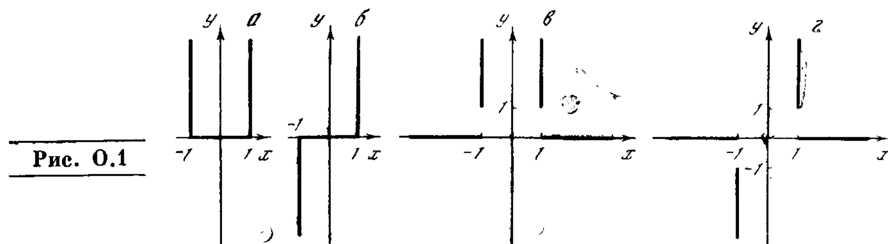


Рис. О.1

в точку $M_1(x_1,\ y_1)$, где $x_1=kx$, $y_1=y$ $\left(\text{откуда } x=\frac{1}{k}x_1,\ y=y_1\right)$. **1.4.4.** Требуется доказать, что при $k,\ x_1,\ x_2>0$ имеет место неравенство $\frac{1}{2}\left(\frac{k}{x_1}+\frac{k}{x_2}\right)>\frac{k}{(x_1+x_2)/2}$, или $\frac{x_1+x_2}{2x_1x_2}>\frac{2}{x_1+x_2}$.

1.5.1. в) График функции $y=x^4+x^2$ получается «сложением» известных нам графиков функций $y_1=x^4$ и $y_2=x^2$; по форме он похож на эти графики. 1.5.3. См. рис. О.1, *а*—*г*, отвечающие предельным случаям $n\to\infty$; при больших n графики будут близки к изображенным на рис. О.1.

1.6.1. а) $y=\frac{x}{2}-2$; б) $y=\pm\sqrt{x+2}$; в) $y=\sqrt[3]{x+1}-1$; г) $y=\pm\sqrt{\sqrt{x+1}-1}$.

1.6.3. а) $y=\frac{x}{a}-\frac{b}{a}$; б) $y=\pm\sqrt{\frac{x-(c-b^2/4a)}{a}}-\frac{b}{2a}$; в) $y=\frac{dx-b}{-cx+a}$.

1.7.8. Здесь не сложно и чисто формальное доказательство того, что фигурирующие в условии преобразования переменных x и y сохраняют знак величины aB (см. решение упр. 9), откуда уже следует несводимость кривых (10а), (10б) и (10в) одна к другой. Однако возможно и более содержательное («геометрическое») рассуждение. Нетрудно понять, что ни сдвиги начала координат, ни изменения масштабов по осям x и y, равносильные равномерным растяжениям кривой в направлениях осей, не сказываются на наличии или отсутствии у кривой $y=f(x)$ (локальных) максимумов и минимумов, а также на наличии или отсутствии у нее точек перегиба, касательная в которых горизонтальна. Отсюда и из того, что кривая (10в) имеет два максимума (отсутствующие у монотонно возрастающих кривых (10а) и (10б)), а кривая (10а) имеет точку перегиба (касательная в которой горизонтальна) — начало координат, отсутствующую у двух других кривых, уже следует требуемое утверждение. **1.7.9.** При $B=c-\frac{b^2}{3a}=0$ зависимость (9) сводится к виду (10а); если же $B\neq 0$, то эта зависимость сводится к виду (10б) при $aB=ac-\frac{b^2}{3}>0$ и к виду (10в) при $ac-\frac{b^2}{3}<0$.

1.8.1. а) Кривая изображена на рис. О.2. **1.8.2.** Окружность $x^2+y^2=1$. **1.8.3.** Эллипс (рис. О.3). **1.8.4.** Отрезок прямой $y=x$ между точками $(-1, -1)$ и $(1, 1)$. **1.8.5.** $x=r\arccos\frac{r-y}{r}-\sqrt{2ry-y^2}$. **1.8.6.** а) Уравнения эпициклоиды: $x=(R+r)\times\cos t-r\cos\frac{R+r}{r}t$, $y=(R+r)\sin t-r\sin\frac{R+r}{r}t$. б) Уравнения гипоциклоиды:

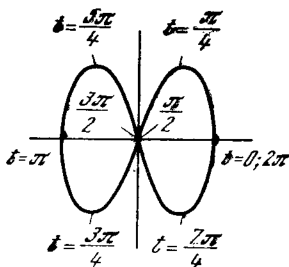


Рис. О.2

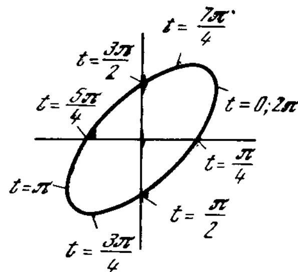


Рис. О.3

$x=(R-r)\cos t+r\cos\frac{R-r}{r}t$, $y=(R-r)\sin t-r\sin\frac{R-r}{r}t$. На рис. О.4, *а—г* изображены соответственно: эпициклоида при $R=r$ (кардиода); гипоциклоида при $r={}^1/_3R$ (кривая Штейнера); гипоциклоида при $r={}^1/_4R$ (астроида); гипоциклоида при $r={}^1/_2R$ (диаметр окружности радиуса R).

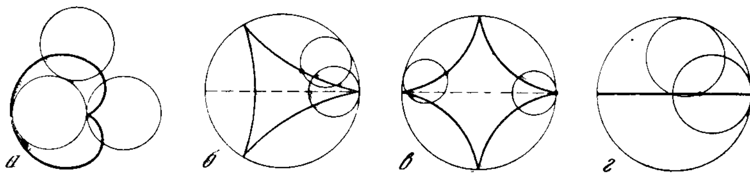


Рис. О.4

1.9.1. а) $y=-x$; б) $x=0$; в) $y=x+2$; г) $y=-{}^4/_3x$. **1.9.2.** Для первой тройки точек: $\widehat{A_2A_1A_3}=90°$; $S_{\triangle A_1A_2A_3}=25$; $h_{A_1,\,A_2A_3}=2\sqrt{5}$. **1.9.3.** а) $2\sqrt{2}$. **1.9.4.** а) $3\sqrt{2}/2$. **1.9.7.** Если $y=kx+s$ — уравнение прямой, то в «новых» координатах x', y', связанных со «старыми» координатами x, y формулами (6), уравнение той же прямой будет иметь вид $y'=k'x'+s'$, где

$$k'=\frac{k\cos\alpha'+\sin\alpha'}{\cos\alpha'-k\sin\alpha'},\qquad s'=\frac{s+ka'-b'}{\cos\alpha'-k\sin\alpha'};$$

здесь $\alpha'=-\alpha=\angle x'Ox$ — угол между осями $O'x'$ и Ox, а a' и b' — координаты нового начала O' в старой системе координат xOy.

Глава 2

2.1.1. а) $v_{\text{ср}}=3t^2+1+3t\Delta t+(\Delta t)^2$; $v_{\text{мгн}}=3t^2+1$. **2.1.2.** $v_{\text{ср}}=v_{\text{мгн}}$.

2.2.1. $c_{\text{ср}}=0{,}3965+4{,}162\cdot10^{-3}T-15{,}072\cdot10^{-7}T^2+(2{,}081\cdot10^{-3}-15{,}072\cdot10^{-7}T)\times\Delta T-5{,}024\cdot10^{-7}(\Delta T)^2$; $c_{\text{мгн}}=0{,}3965+4{,}162\cdot10^{-3}T-15{,}072\cdot10^{-7}T^2$; $c(0)=0{,}3965$; $c(100)=0{,}7977$; $c(500)=2{,}1007$. **2.2.2.** $c_{\text{мгн}}=4186{,}68+16746{,}72\cdot10^{-5}T+3768\cdot10^{-6}T^2$.

2.4.1. а) $y'=4x^3$; б) $y'=12x^2-6x+2$; в) $y'=8x+4$; г) $y'=-2/x^3$; д) $y'=a(1-1/x^2)$; е) $y'=2ax-2b/x^3$. **2.4.4.** $(1,2)^2=1{,}44$; если $y(x)=x^2$, $x=1$, $\Delta x=0{,}2$, то $y(x)+y'(x)\Delta x=1^2+2\cdot0{,}2=1{,}4$; ошибка составляет около 3%. Аналогично $(1{,}1)^2=1{,}21$ и $(1{,}1)^2\approx1^2+2\cdot0{,}1=1{,}2$ (ошибка около 1%) и т. д.

2.5.2. Касательная горизонтальна в точках $x=0$ и $x=2$. **2.5.3.** См. рис. О.5; горизонтальная касательная при $x=\pm1/\sqrt{3}$. **2.5.4.** См. рис. О.6. **2.5.5.** См. рис. О.7. **2.5.6.** а) $y={}^3/_4\,x-{}^1/_4$, $({}^1/_3, 0)$, $(0, -{}^1/_4)$; б) $y=3x-2$, $({}^2/_3, 0)$, $(0, -2)$. **2.5.7.** Касательная к параболе а) в точке (x_0, y_0) пересекает оси координат в точках $(x_0/2, 0)$ и $(0, -y_0)$; касательная к кубической параболе б) пересекает оси координат в точках $({}^2/_3\,x_0, 0)$, $(0, -2y_0)$.

2.6.1. а) $x=0$; минимум при $a>0$, максимум при $a<0$; б) $x=-1$ — максимум; $x=1$ — минимум; в) $x=-\sqrt{b/a}$ $(b/a>0)$, максимум при $a>0$, минимум при $a<0$; $x=\sqrt{b/a}$ $(b/a>0)$, максимум при $a<0$, минимум при $a>0$. При $b/a\leqslant0$ нет ни мак-

симума, ни минимума; г) $x=-1$ — максимум; $x=1$ — минимум; е) $a>0$, $x=0$ — минимум; $a=0$, $x=0$ — минимум; $a<0$, $x=0$ — максимум, $x=-\sqrt{-a/2}$ — минимум, $x=\sqrt{-a/2}$ — минимум. **2.6.2.** $t=4°$. **2.6.3.** а) $t\approx 4{,}08°$, б) $t\approx 3{,}92°$.

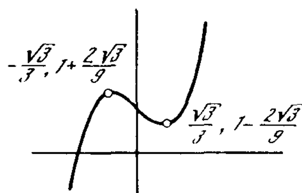


Рис. О.5

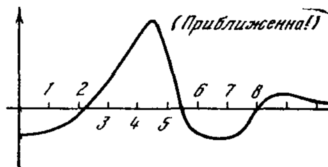


Рис. О.6

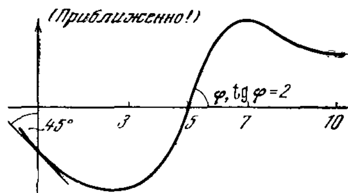


Рис. О.7

2.7.1. 2; $6x$; $12x^2$; $2a$. **2.7.2.** Ускорение равно $2a$; оно постоянно. **2.7.3.** а, в) Кривая выпукла при $x<0$, вогнута при $x>0$; б) выпукла при $x<-p/3$, вогнута при $x>-p/3$.

Глава 3

3.2.2. Ниже приведены значения сумм при разбиении отрезка на m частей ($m=10$, 20, 50, ∞):

m	10	20	50	∞
Δt	0,1	0,05	0,02	0
$\Sigma t^2_{l-1}\cdot\Delta t$	2,18	2,26	2,30	2,33
$\Sigma t^2_l\cdot\Delta t$	2,49	2,41	2,37	2,33

Видно, что уже при $m=50$ обе суммы незначительно отличаются от предельного значения при $m=\infty$.

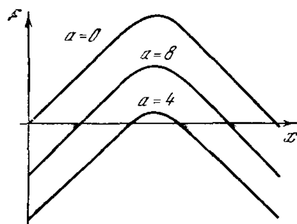


Рис. О.8

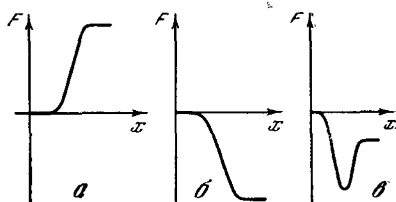


Рис. О.9

Для точного вычисления интеграла (предельной суммы, отвечающей «значению» $m=\infty$) запишем:

$$\sum_l t_l^2\frac{1}{m}=\sum_{l=1}^{m}\left(\frac{l}{m}\right)^2\frac{1}{m}=(1^2+2^2+\ldots+m^2)\frac{1}{m^3}=\frac{m(m+1)(2m+1)}{6m^3}=$$
$$=\frac{2m^2+3m+1}{6m^2}=\frac{1}{3}+\frac{1}{2}\,\frac{1}{m}+\frac{1}{6}\,\frac{1}{m^2},$$

откуда сразу следует, что $\int\limits_1^2 t^2dt=\frac{1}{3}\ (\approx 0{,}33)$. **3.2.3.** Пусть $q=\sqrt[m]{2}$; обозначим $x_0=1$, $x_1=q$, $x_2=q^2$, ..., $x_m=q^m=2$. Далее составим $\sum\limits_l x_l^3\Delta x_l=\sum\limits_l (x_l)^3(x_l-x_{l-1})=$

$$=\sum_l (q^l)^3(q^l-q^{l-1})=(q-1)\sum_{l=1}^{m}q^{4l-1}=(q-1)(q^3+q^7+q^{11}+\ldots+q^{4m-1})=(q-1)\times$$

$\times \frac{q^{4m+3}-q^3}{q^4-1}=\frac{q^3[(q^m)^4-1]}{q^3+q^2+q+1}=\frac{q^3(2^4-1)}{q^3+q^2+q+1}$; устремляя $m\to\infty$ (и $q\to 1$), получим $\int\limits_1^2 x^3dx=\frac{15}{4}=3\frac{3}{4}=3{,}75$.

3.4.1. а) 1/3; б) $\frac{1}{3}(1{,}1^3-1)\approx 0{,}11033$; в) 1/2; г) $2(\sqrt{3}-1)\approx 1{,}464$. **3.4.2.** а) $S=\int\limits_0^b y(x)\,dx$, где $y=y(x)=\frac{h}{b}x$ — уравнение гипотенузы. Отсюда $S=\int\limits_0^b \frac{h}{b}xdx=\frac{h}{b}\int\limits_0^b xdx=\frac{h}{b}\frac{x^2}{2}\Big|_0^b=\frac{1}{2}bh$. **3.4.3.** а) $S={}^1/_3x_0y_0$; б) $S={}^2/_3x_0y_0$. **3.4.4.** $S=\int\limits_{-r}^r\sqrt{r^2-x^2}dx$. **3.4.5.** 0,7837 для $n=5$; 0,7850 для $n=10$. **3.4.6.** См. рис. О.8. **3.4.7.** См. рис. О.9, *а—в*.

3.6.2. Не изменится. **3.6.3.** ${}^2/_3R^3\operatorname{tg}\alpha$. **3.6.4.** $\int\limits_0^1 y\sqrt{1-y^2}dy=1/3$; заметьте, что площадь сечения «копыта» (рис. 3.6.8) плоскостью, перпендикулярной плоскости ABP, параллельной AB и удаленной от AB на расстояние y, равна $2y\sqrt{1-y^2}$.

Глава 4

4.2.1. а) $y'=(x^2)'x^2+x^2(x^2)'=2\cdot 2x\cdot x^2=4x^3$; б) $y'=5ax^4+4bx^3+3cx^2+2dx+e$; г) $(5x+3/2)\sqrt{x}$. **4.2.2.** $f'''(x)=g(x)h'''(x)+3g'(x)h''(x)+3g''(x)h'(x)+g'''(x)h(x)$. **4.2.3.** $f''(x)=ghk''+gh''k+g''hk+2gh'k'+2g'hk'+2g'h'k$.

4.3.1. $\frac{dz}{dx}=\frac{dz}{dy}\frac{dy}{dx}=2ya=2a(ax+b)$; $z'=(ax^2+2abx+b^2)'=2a^2x+2ab$. **4.3.2.** а) $z'=-\frac{a}{(ax+b)^2}$; б) $z'=-\frac{2a}{(ax+b)^3}$; в) $z'=\frac{1}{(x+1)^2}$; г) $y'=\frac{2x(x^2+4x+5)}{(x+1)^2}$; д) $y'=-\frac{x^2-2x-2}{(x^2+2)^2}$. **4.3.3.** $y''=\frac{f''g^2-2f'gg'+2f(g')^2-fgg''}{g^3}$.

4.4.1. а) $y'=\frac{2}{3}\frac{1}{\sqrt[3]{(2x+1)^2}}$; б) $y'=\frac{1}{(x+1)\sqrt{x^2-1}}$; в) $y'=-\frac{1}{4}\frac{1}{x\sqrt[4]{x}}$. **4.4.2.** $\frac{dy}{dx}=-\frac{b}{a}\operatorname{ctg}t$. **4.4.3.** б) $\frac{d^2y}{dx^2}=\frac{5}{16}\frac{1}{x^2\sqrt[4]{x}}$; в) $\frac{d^2y}{dx^2}=-\frac{b}{a^2}\frac{1}{\sin^3 t}$.

4.5.1. а) $y'=5x^4-12x^3+3x^2+14x-2$; б) $y'=2(x^3+x+1)(3x^2+1)$; в) $y'=4(x^2-x+1)^3(2x-1)$; г) $y'=10(3x^2-1)^9\cdot 6x$; д) $y'=\frac{1}{\sqrt{x^2-1}}$; е) $y=x^{2/5}$, $y'=\frac{2}{5}\frac{1}{\sqrt[5]{x^3}}$. **4.5.3.** а) Если x изменяется на 1%, то $\Delta y=n\cdot 0{,}01y$; поэтому при изменении x на k % $\Delta y=n\cdot 0{,}01yk$. В данном случае x изменяется на 10%, $\Delta y=n\cdot 0{,}1y$. Так как $n=1/2$, то $\Delta y=\frac{1}{2}\cdot 0{,}1y=0{,}05y$. Поэтому $y(11)\approx y(10)+0{,}05\cdot 5=5{,}25$; $y(9)\approx y(10)-0{,}05\cdot 5=4{,}75$.

Получим теперь точное решение. Обозначим коэффициент пропорциональности через k; тогда $y=k\sqrt{x}$. Так как при $x=10$ должно быть $y=5$, то $5=k\sqrt{10}$, откуда $k\approx 1{,}58$ (вычисления ведем с двумя знаками после запятой). Поэтому $y=1{,}58\sqrt{x}$; $y(11)=1{,}58\sqrt{11}\approx 5{,}24$; $y(9)\approx 4{,}74$. б) Приближенные значения: $y(11)\approx 4{,}50$; $y(9)\approx 5{,}50$. Точные значения: $y(11)\approx 4{,}54$; $y(9)\approx 5{,}56$. в) Приближенные значения: $y(11)\approx 6{,}00$; $y(9)\approx 4{,}00$. Точные значения: $y(11)\approx 6{,}05$; $y(9)\approx 4{,}05$.

4.6.1. $y'=x^2(x^2-1)(7x^2-3)$. **4.6.2.** $y'=3x^2\sqrt{x^2+x}+\frac{(2x+1)x^3}{2\sqrt{x^2+x}}\left(=\frac{(8x+7)x^3}{2\sqrt{x^2+x}}\right)$.

4.6.3. $y'=5x^4\sqrt[3]{x^2-1}(x^3-2x)^{1/5}+\frac{1}{3}\frac{\sqrt[3]{x^2-1}}{x^2-1}\cdot 2xx^5(x^3-2x)^{1/5}+\frac{1}{5}\frac{(x^3-2x)^{1/5}}{x^3-2x}(3x^2-2)\times$

$\times\, x^5\sqrt[3]{x^2-1}$ (упростите!). **4.6.4.** $y'=\left(1-\dfrac{1}{2\sqrt{x^3}}\right)\sqrt{x^3-2}+\left(x+\dfrac{1}{\sqrt{x}}\right)\dfrac{3x^2}{2\sqrt{x^3-2}}$. **4.6.5.** $y'=2x\sqrt{\sqrt[3]{x}+x}+x^2\dfrac{\dfrac{1}{3\sqrt[3]{x^2}}+1}{2\sqrt{\sqrt[3]{x}+x}}$ (упростите!). **4.6.6.** $y'=5\left(\sqrt[3]{x}+\dfrac{1}{\sqrt[3]{x}}\right)^4\left(\dfrac{1}{3\sqrt[3]{x^2}}-\dfrac{1}{3\sqrt[3]{x^4}}\right)x+\left(\sqrt[3]{x}+\dfrac{1}{\sqrt[3]{x}}\right)^5$. **4.6.7.** $y'=\dfrac{x^2+1}{(x^2-1)^2}$. **4.6.8.** $y'=2\dfrac{1-x^2}{(x^2-x+1)^2}$. **4.6.9.** $y'=\dfrac{x^2+2x+5}{(x+1)^2}$. **4.6.10.** $y'=\dfrac{-12x+5}{x^6}\sqrt{x^3+2}+\dfrac{3(3x-1)}{2x^2\sqrt{x^3+2}}$. **4.6.11.** $y'=-\dfrac{1}{(x^2-1)^{3/2}}$. **4.6.12.** $y'=\dfrac{2x+3}{3(x+1)^{4/3}}$. **4.6.13.** $y'=\dfrac{4x+3\sqrt{x}}{4\sqrt{x^2+x\sqrt{x}}}$. **4.6.14.** $y'=\dfrac{6x\sqrt[3]{x^2}+1}{6\sqrt[3]{x^2}\sqrt{x^2+\sqrt[3]{x}}}$. **4.6.15.** $y'=\dfrac{(8x-x^3)\sqrt{1+x^2}}{(1+x^2)^3}$. **4.6.16.** $y'=\sqrt[3]{(2x+3)^2}+x\dfrac{4}{3\sqrt[3]{2x+3}}$. **4.6.17.** $y'=3x^2\sqrt{x-1}+\dfrac{x^3-1}{2\sqrt{x-1}}+\sqrt[3]{x^2-1}+\dfrac{2x^2}{3\sqrt[3]{(x^2-1)^2}}$. **4.6.18.** $y'=\dfrac{-2x^2-x+9}{3(x-1)^3\sqrt{2x-3}}$. **4.6.19.** $y'=\sqrt{\dfrac{x+1}{x-1}}\dfrac{1}{(x+1)^2}$. **4.6.20.** $y'=\dfrac{4x^3-10x^2-22x-11}{3(x-2)^2\sqrt[3]{(x+1)^2}}$. **4.6.21.** $y'=\dfrac{-x^5+2x^3+2x^2-1}{(x^3+1)^2\sqrt{x^2-1}}$. **4.6.22.** $y'=\dfrac{1}{3}\left(\dfrac{x+1}{x^2+x+1}\right)^{2/3}\dfrac{x^2+2x}{(x+1)^2}$. **4.6.23.** $y'=\sqrt{x^2-1}\sqrt[3]{x+\sqrt{x}}+\dfrac{x^2}{\sqrt{x^2-1}}\sqrt[3]{x+\sqrt{x}}+\dfrac{(1+2\sqrt{x})x\sqrt{x^2-1}}{6\sqrt{x}\sqrt[3]{(x+\sqrt{x})^2}}$. **4.6.24.** $y'=\dfrac{1}{7}\left(x+\dfrac{1}{\sqrt{x}}\right)^{-6/7}\left(1-\dfrac{1}{2\sqrt{x^3}}\right)x^2+2x\left(x+\dfrac{1}{\sqrt{x}}\right)^{1/7}$. **4.6.25.** $y'=\dfrac{1-6\sqrt[3]{x^2}-4x\sqrt[3]{x^2}}{3(x+1)\sqrt[3]{x^2(x+1)}}$.

4.7.1. $y'\approx 2{,}3\lg 2\cdot 2^x$. **4.7.2.** $y'\approx 2{,}3\lg 5\cdot 5^{x+1}$. **4.7.3.** $y'\approx -2{,}3\lg 2\cdot\left(\dfrac{1}{2}\right)^x$. **4.7.4.** $y'\approx 2{,}3\cdot 10^{\sqrt{x}}\dfrac{1}{2\sqrt{x}}$. **4.7.5.** $y'\approx 2x\cdot 2{,}3\lg 2\cdot 2^{x^2}$. **4.7.6.** $y'\approx\left(1-\dfrac{1}{x^2}\right)\cdot 2{,}3\lg 2\cdot 2^{x+1/x}$.

4.8.1. а) $y'=-e^{-x}$; б) $y'=5e^x-3e^{3x}$; в) $y'=2xe^{x^2}$; г) $y'=\dfrac{1}{2\sqrt{x}}e^{\sqrt{x}}$; д) $y'=3(x^2-1)e^{x^3-3x+1}$. **4.8.2.** Больше чем в 150 000 раз.

4.9.1. $\log_5 15=\dfrac{\lg 15}{\lg 5}\approx 1{,}6825$. **4.9.2.** а) Дифференцируя обе части указанного равенства (логарифмы — натуральные!), находим $\dfrac{(uv)'}{uv}=\dfrac{u'}{u}+\dfrac{v'}{v}$, откуда $(uv)'=u'v+uv'$. б) Продифференцируйте почленно равенство $\ln(u/v)=\ln u-\ln v$. **4.9.3.** а) $y'=\dfrac{1}{x+3}$; б) $y'=\dfrac{1}{2x}(2x)'=\dfrac{1}{x}$, или $y=\ln 2x=\ln 2+\ln x$, поэтому $y'=(\ln 2)'+(\ln x)'=\dfrac{1}{x}$; в) $y'=\dfrac{2x}{x^2+1}$; г) $y'=\dfrac{x^2-1}{x(x^2+1)}$; д) $y'=\dfrac{6x-1}{3x^2-x+1}$; е) $y'=\dfrac{2}{x^2-1}$; ж) $y'=\dfrac{x+3}{6x(x+1)}$; з) $y'=\ln x+1$; и) $y'=3x^2\ln(x+1)+\dfrac{x^3}{x+1}$; к) Так как $\ln y=x\ln x$, то, дифференцируя, находим $\dfrac{1}{y}y'=\ln x+1$, откуда $y'=y(\ln x+1)$, или $y'=x^x(\ln x+1)$; л) $y'=\dfrac{1}{2}(\sqrt{x})^{\sqrt{x^2-1}}\left(\dfrac{x\ln x}{\sqrt{x^2-1}}+\dfrac{\sqrt{x^2-1}}{x}\right)$ (см. указание к упр. к).

4.10.1. $y'=2\cos(2x+3)$. **4.10.2.** $y'=-\sin(x-1)$. **4.10.3.** $y'=-(2x-1)\sin(x^2-x+1)$. **4.10.4.** $y'=2\sin x\cos x$. **4.10.5.** $y'=3\cos 3x\cos^2 x-2\cos x\sin x\sin 3x$. **4.10.6.** $y'=(\sin 2x)^x\left(\ln\sin 2x+\dfrac{2x\cos 2x}{\sin 2x}\right)$ (ср. с упр. 9.3к); **4.10.7.** $y'=\operatorname{tg}x+\dfrac{x}{\cos^2 x}$. **4.10.8.** $y'=\dfrac{2}{\cos^2 2x}e^{\operatorname{tg}2x}$. **4.10.9.** $y'=-\dfrac{1}{2}\dfrac{1}{\sin^2(x/2)}$.

4.11.1. $y' = -\frac{1}{\sqrt{1-x^2}}$. 4.11.2. $y' = -\frac{1}{1+x^2}$. 4.11.3. $y' = \frac{2}{\sqrt{1-4x^2}}$. 4.11.4. $y' = \frac{3}{9x^2+6x+2}$. 4.11.5. $y' = \frac{2x-1}{x^4-2x^3+x^2+1}$. 4.11.6. $y' = \frac{1}{2\sqrt{x}(x+1)} e^{\operatorname{arc\,tg}\sqrt{x}}$.

4.12.1. -1; 1. 4.12.2. -1.

Глава 5

5.2.1. 1/6. 5.2.2. 1/2. 5.2.3. 2. 5.2.4. $\frac{\pi}{4}\left(=\operatorname{arc\,tg} 1 - \operatorname{arc\,tg} 0 = \frac{\pi}{4} - 0\right)$. 5.2.5. $e - \frac{1}{e}$. 5.2.6. 1.

5.3.1. $x^3 - x^2 + x + C$. 5.3.2. $\frac{4}{5}x^5 - \frac{3}{4}x^4 + \frac{x^3}{3} - \frac{x^2}{2} + C$. 5.3.3. $\frac{1}{4}x^4 - \frac{2}{3}x^3 + \frac{1}{2}x^2 + C$. 5.3.4. $\frac{1}{2}x^2 + 2x + 3\ln|x| + C$. 5.3.5. $2x + \ln|x-1| + C$. 5.3.6. $\frac{a}{c}x + \frac{bc-ad}{c^2}\ln|cx+d| + C$. 5.3.7. Если $\frac{x}{(x-2)(x-3)} = \frac{A}{x-2} + \frac{B}{x-3}$, то $A(x-3) + B(x-2) = x$, т. е. $(A+B)x - (3A+2B) = x$ и $A + B = 1$, $3A + 2B = 0$; поэтому $A = -2$, $B = 3$. Ответ. $-2\ln|x-2| + 3\ln|x-3| + C$. 5.3.8. $-\ln|x-1| + \ln|x-2| + C$. 5.3.9. Если $\frac{x+1}{x^2-3x+2} = \frac{x+1}{(x-1)(x-2)} = \frac{A}{x-1} + \frac{B}{x-2}$, то $A(x-2) + B(x-1) = x+1$, откуда $A = -2$, $B = 3$ (положите в предыдущем равенстве сначала $x = 1$, а затем $x = 2$). Ответ. $-2\ln|x-1| + 3\ln|x-2| + C$. 5.3.10. Если $\frac{x+2}{x(x^2+1)} = \frac{A}{x} + \frac{B}{x^2+1} + \frac{Cx}{x^2+1}$, то $A(x^2+1) + Bx + Cx^2 = x + 2$, т. е. $A = 2$, $B = 1$, $C = -2$. Ответ. $2\ln|x| + \operatorname{arc\,tg} x - \ln(x^2+1) + C$.

5.4.1. $\cos x + x\sin x + C$. 5.4.2. $x(\ln|x| - 1) + C$. [Указание. Положите в формуле (5.4.3) $f = \ln x$, $dg = dx$.] 5.4.3. $\frac{1}{2}x\sin 2x - \left(\frac{1}{2}x^2 - \frac{1}{4}\right)\cos 2x + C$. 5.4.4. $(-x^3 - 3x^2 - 6x - 6)e^{-x} + C$. 5.4.5. $(2x+1)\cos x + (x^2 + x - 1)\sin x + C$. 5.4.6. Пусть $(2x^2+1)\cos 3x\,dx = (a_1x^2 + b_1x + c_1)\cos 3x + (a_2x^2 + b_2x + c_2)\sin 3x$. Дифференцируя обе части равенства, получаем $(2x^2+1)\cos 3x = (2a_1x + b_1)\cos 3x - (3a_1x^2 + 3b_1x + 3c_1) \times \sin 3x + (2a_2x + b_2)\sin 3x + (3a_2x^2 + 3b_2x + 3c_2)\cos 3x$, т. е. $(2x^2+1)\cos 3x = (-3a_1x^2 - 3b_1x - 3c_1 + 2a_2x + b_2)\sin 3x + (3a_2x^2 + 3b_2x + 3c_2 + 2a_1x + b_1)\cos 3x$. Таким образом, имеем $2x^2 + 1 = 3a_2x^2 + 3b_2x + 3c_2 + 2a_1x + b_1$, $0 = -3a_1x^2 - 3b_1x - 3c_1 + 2a_2x + b_2$, откуда $3a_2 = 2$, $3b_2 + 2a_1 = 0$, $3c_2 + b_1 = 1$, $-3a_1 = 0$, $-3b_1 + 2a_2 = 0$, $-3c_1 + b_2 = 0$. Из этой системы уравнений находим: $a_1 = 0$, $b_2 = 0$, $c_1 = 0$, $a_2 = \frac{2}{3}$, $b_1 = \frac{4}{9}$, $c_2 = \frac{5}{27}$. Ответ. $\frac{4}{9}x\cos 3x + \left(\frac{2}{3}x^2 + \frac{5}{27}\right)\sin 3x + C$. 5.4.7. $x\arcsin x + \sqrt{1-x^2} + C$. 5.4.8. $x\operatorname{arc\,tg} x - \frac{1}{2}\ln(x^2+1) + C$. 5.4.9. Положим $f = \sin 3x$, $dg = e^{2x}dx$; тогда интегрирование по частям даст $\int e^{2x}\sin 3x\,dx = \frac{1}{2}e^{2x}\sin 3x - \frac{3}{2}\int e^{2x}\cos 3x\,dx$. В последнем интеграле положим $f = \cos 3x$, $dg = e^{2x}dx$ и снова выполним интегрирование по частям; мы получим $\int e^{2x}\sin 3x\,dx = \frac{1}{2}e^{2x}\sin 3x - \frac{3}{2}\left(\frac{1}{2}e^{2x}\cos 3x + \frac{3}{2}\int e^{2x}\sin 3x\,dx\right)$. Рассматривая последнее равенство как уравнение с неизвестным $\int e^{2x}\sin 3x\,dx$, находим $\int e^{2x}\sin 3x\,dx = \frac{e^{2x}(2\sin 3x - 3\cos 3x)}{13} + C$. 5.4.10. $\frac{e^x(\cos 2x + 2\sin 2x)}{5} + C$.

5.5.1. $\frac{1}{3}\sin(3x-5) + C$ $\left(\text{положите } 3x - 5 = t,\ dx = \frac{dt}{3}\right)$. 5.5.2. $-\frac{1}{2}\cos(2x+1) + C$. 5.5.3. $\frac{2}{9}\sqrt{(3x-2)^3} + C$. 5.5.4. Так как при $\sqrt{x} = z$ имеем $x = z^2$ и $dx = 2z\,dz$, то $\int\frac{x\,dx}{x+\sqrt{x}} = \int\frac{z^2\cdot 2z\,dz}{z^2+z} = 2\int\frac{z^2\,dz}{1+z} = 2\int\frac{z^2-1+1}{1+z}dz = 2\int\left[(z-1) + \frac{1}{1+z}\right]dz = z^2 - 2z + 2\ln(1+z) + C = x - 2\sqrt{x} + 2\ln(1+\sqrt{x}) + C$. 5.5.5. $\sqrt{x^2-5} + C$. 5.5.6. $\frac{\sin^4 x}{4} + C$.

или $-\frac{1}{2}\cos^2 x+\frac{1}{4}\cos^4 x+C$. 5.5.7. $-\frac{1}{3\sin^3 x}+\frac{1}{\sin x}+C$. 5.5.8. $-\ln|\cos x|+C$ (положите $\cos x=t$). 5.5.9. $\arcsin\left(\frac{x}{a}\right)+C$. 5.5.11. $\ln\left|x+\sqrt{a^2+x^2}\right|+C$. (Помимо намеченного на с. 145 приема, здесь возможна также искусственная подстановка $x=a\operatorname{tg} t$ (т. е. $t=\operatorname{arctg}(x/a)$), $dx=a\,dt/\cos^2 t$ и $a^2+x^2=a^2/\cos^2 t$ или подстановка $\sqrt{a^2+x^2}=z-x$, откуда, возводя обе части последнего равенства в квадрат, получаем $a^2=z^2-2zx$ (член с x^2 в обеих частях равенства сокращается!) и, значит, $x=\frac{z^2-a^2}{2z}$, $\sqrt{a^2+x^2}=z-x=z-\frac{z^2-a^2}{2z}=\frac{z^2+a^2}{2z}$, $dx=\frac{1}{2}\frac{z\cdot 2z-(z^2-a^2)}{z^2}dz=$ $=\frac{z^2+a^2}{2z^2}dz$, откуда $\int\frac{dx}{\sqrt{a^2+x^2}}=\int\left[\frac{(z^2+a^2)}{2z^2}:\frac{z^2+a^2}{2z}\right]dz=\int\frac{dz}{z}$.) 5.5.12. Положим $x=a\operatorname{tg} t$; тогда $x^2+a^2=\frac{a^2}{\cos^2 t}$, $dx=\frac{a\,dt}{\cos^2 t}$, так что $\int\frac{dx}{(x^2+a^2)^2}=\frac{1}{a^3}\int\cos^2 t\,dt=$ $=\frac{1}{2a^3}\int(1+\cos 2t)\,dt=\frac{1}{2a^3}\left(t+\frac{1}{2}\sin 2t\right)+C$. А так как $t=\operatorname{arctg}\frac{x}{a}$ и $\sin 2t=$ $=\frac{2\operatorname{tg} t}{1+\operatorname{tg}^2 t}=\frac{2x/a}{1+(x/a)^2}=\frac{2ax}{x^2+a^2}$, то окончательно имеем $\int\frac{dx}{(x^2+a^2)^2}=\frac{1}{2a^3}\times$ $\times\operatorname{arctg}\frac{x}{a}+\frac{1}{2a^2}\frac{x}{x^2+a^2}+C$.

Глава 6

6.1.1. $y=(ax_0^3+bx_0^2+cx_0+d)+(3ax_0^2+2bx_0+c)(x-x_0)+(3ax_0+b)(x-x_0)^2+$ $+a(x-x_0)^3$. Последующие члены все равны нулю; сумма выписанных четырех членов равна многочлену. 6.1.2. Так как $y(0)=0$, $y'(0)=1$, $y''(0)=2$, ..., $y^{(n)}(0)=n$, ..., то $y=x+x^2+\frac{x^3}{2!}+\ldots=x\left(1+x+\frac{x^2}{2!}+\ldots\right)$. 6.1.3. $y=e\left[1+(x-1)+\frac{1}{2!}(x-1)^2+\right.$ $\left.+\frac{1}{3!}(x-1)^3+\ldots\right]$. 6.1.4. Легко проверить, что $(1+r)^m\approx 1+mr+\frac{m(m-1)}{2}r^2$ (ср. § 4), $e^{mr}\approx 1+mr+\frac{m^2r^2}{2}$. Таким образом, при малом r величина e^{mr} отличается от $(1+r)^m$ слагаемым $^1/_2 mr^2$. Так, если $m=50$, $r=0{,}02$ (см. пример на с. 127), то $^1/_2 mr^2=0{,}01$, т. е. ошибка менее 0,5%. (При малом mr^2 величина m может быть и большой; при этом, конечно, и величины mr^3, mr^4, ... будут малыми.) 6.1.5. а) По формуле (24), где $y(0)=1$ и $\Delta y=y(\Delta x)-1$, имеем

Δx	1	1/2	1/4	1/8	$\Delta x\to 0$
$y(\Delta x)$	2,7183	1,6487	1,2840	1,1331	1
$\frac{\Delta y}{\Delta x}$	1,718	1,297	1,136	1,065	1

По формуле (25), полагая $\Delta y=y\left(\frac{\Delta x}{2}\right)-y\left(-\frac{\Delta x}{2}\right)$, получаем

Δx	1	1/2	1/4	1/8	$\Delta x\to 0$
$y\left(\frac{\Delta x}{2}\right)$	1,6487	1,2840	1,1331	1,0645	1
$y\left(-\frac{\Delta x}{2}\right)$	0,6065	0,7788	0,8825	0,9394	1
Δy	1,0422	0,5052	0,2506	0,1251	Δx
$\frac{\Delta y}{\Delta x}$	1,042	1,010	1,002	1,006	1

6.1.6. Подставьте в числитель дроби в правой части (27) выражение (13а) для $f(a+\Delta x)$ и $f(a-\Delta x)$ (в (13а) надо положить соответственно $x=a+\Delta x$ и $x=a-\Delta x$).

6.2.1. Примените формулу Тейлора (1.18) к функции (1) (или формулу Маклорена (1.19) к функции (1а). **6.2.2.** $\mathrm{si}\, x = C + x - \frac{x^3}{18} + \frac{x^5}{600} - \ldots$, где $C=0$, если условиться, что $\mathrm{si}\, 0 = 0$.

6.3.1. а) $y = \frac{x+1}{1-x} = 1 + 2x + 2x^2 + 2x^3 + 2x^4 + \ldots$; б) $y = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots$ **6.3.2.** $y = \ln x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \frac{(x-1)^4}{4} + \ldots$ В упр. 1 ряды пригодны для вычислений при $|x|<1$; в упр. 2 — при $0<x<2$.

6.3.3. $f(x)g(x) = f(0)g(0) + [f'(0)g(0) + g'(0)f(0)]x + \frac{1}{2}[f''(0)g(0) + 2f'(0)g'(0) + g''(0)f(0)]x^2 + \ldots$

6.4.2. Это суть формулы (3) для $a=1$ и $m=\frac{1}{n}$, в которых оставлены соответственно первые два (а) и первые три (б) члена. **6.4.3.** $\sqrt[3]{1,2} \approx 1,067$ и $\approx 1,062$ (табличное значение 1,063); $\sqrt[3]{1,1} \approx 1,033$ и $\approx 1,032$ (табличное значение с четырьмя значащими цифрами: 1,032). **6.4.5.** Уже $(\sqrt{x})'$ при $x=0$ не существует (обращается в ∞).

6.5.1. а) 1; б) —1/2; в) 1/3; г) ∞; д) 1; е) 1/2. **6.5.4.** а) Сделайте подстановку $t=\frac{1}{x}$ и рассмотрите отношение $f\left(\frac{1}{t}\right):g\left(\frac{1}{t}\right)$; воспользуйтесь правилом (2). б) Докажите, что $\frac{f(x)-f(x_0)}{g(x)-g(x_0)} = \frac{f'(c)}{g'(c)}$, где c — некоторое число, промежуточное между x и x_0 (почему?); далее, считая, что x и x_0 близки к a (и расположены по одну сторону от a), устремите сначала $x \to a$, оставляя постоянным x_0 $\left(\text{при этом } \frac{f(x)-f(x_0)}{g(x)-g(x_0)} \to \frac{f(x)}{g(x)}, \text{ где, по условию, } x \to a\right)$, а затем устремите и $x_0 \to a$.

6.6.1. а) $y = -\frac{1}{x+C}$; б) $y^2 - kx^2 = a$; в) $y = Ax^k$; г) $y = \frac{1}{\sqrt{-{}^2/_3 x^3 + C}}$; д) $\mathrm{tg}\left(\frac{y}{2}\right) = Ae^{\sin x}$. **6.6.3.** $y=0$. **6.6.4.** Разложите $e^{x^2/2}\Phi(x)$ в ряд по степеням x (для чего надо разложить в ряд $e^{-t^2/2}$ по степеням t и почленно проинтегрировать полученное равенство); докажите, что при подходящем x_0 произведение $e^{x^2/2}\Phi(x)$ представляется рядом (23) (где $a_0=0$).

6.7.1. $z = z_{\text{ст}} - (z_{\text{ст}} - z_0)\exp\left[-\frac{k}{\pi r_0^2}(t-t_0)\right]$; уровень воды возрастает, асимптотически стремясь к значению $z = z_{\text{ст}}$. **6.7.2.** Найдите удовлетворяющее дифференциальному уравнению стационарное значение $z = z_{\text{ст}} = \mathrm{const}\ (=q_0/k)$; далее обозначьте $z = z_{\text{ст}} + z_1$ и составьте дифференциальное уравнение для функции $z_1 = z_1(t)$.

Глава 7

7.1.1. $x = \frac{a+b-\sqrt{a^2+b^2-ab}}{6}$. **7.1.2.** Если принять за x сторону прямоугольника, принадлежащую основанию треугольника (равному a), то площадь прямоугольника $S(x) = xh\left(1-\frac{x}{a}\right) = hx - \frac{h}{a}x^2$. Решая уравнение $S'(x)=0$, находим $x=\frac{a}{2}$; тогда вторая сторона прямоугольника равна $\frac{h}{2}$ и его площадь $S(x)=\frac{ah}{4}$. **7.1.3.** $S=2R^2$ (искомым прямоугольником является квадрат). **7.1.4.** Радиус основания банки $r=\sqrt[3]{\frac{V}{2\pi}}$, ее высота $h=2r$. **7.1.5.** $t=\frac{av_1+bv_2}{v_1^2+v_2^2}$. **7.1.7.** Время движения $T=\frac{1}{v_1}\sqrt{a^2+x^2}+\frac{1}{v_2}\sqrt{b^2+(c-x)^2}$, где c — величина проекции отрезка AB на линию l раздела сред, a и b — расстояния от точек A и B до l. Условие $\frac{dT}{dx}=0$ дает

$\frac{x}{v_1\sqrt{a^2+x^2}}=\frac{c-x}{v_2\sqrt{b^2+(c+x)^2}}$ или, так как $\frac{x}{\sqrt{a^2+x^2}}=\sin\alpha$, $\frac{c-x}{\sqrt{b^2+(c-x)^2}}=\sin\beta$, где α и β — углы, образуемые линиями движения тела в I и во II средах с перпендикуляром к линии раздела сред, $\frac{\sin\alpha}{\sin\beta}=\frac{v_1}{v_2}$. (Это — *закон Снеллиуса*, т. е. точка должна двигаться так, как движется световой луч, пересекая границу двух сред с разными скоростями движения в этих средах.) Для доказательства того, что мы действительно получаем минимум T, достаточно выписать $\frac{d^2T}{dx^2}$; легко убедиться, что $\frac{d^2T}{dx^2}>0$ при всех x.

7.2.1. $y_{\min}=3$. **7.2.2.** В момент времени t переход $П$ представлен отрезком длины l, концы которого проектируются на прямую, отвечающую берегу, на котором сидит P (мы ее принимаем за ось x), в точках с абсциссами $x_1=v(t-t_0)$, $x_2=v(t-t_0)-l$ (за начало координат принята точка P). Расстояние от P до точки M парохода $П$ равно $\Delta(x)=\sqrt{x^2+h^2}$, где $x_1\leqslant x\leqslant x_2$; нам надо найти $D=\min\Delta(x)$. При $t_0\leqslant t\leqslant t_0+\frac{l}{v}$ этот минимум равен h; он достигается в точке $x=0$, отвечающей условию $\frac{d\Delta}{dx}=0$; при $t<t_0$ и $t>t_0+\frac{l}{v}$ имеют место граничные минимумы, равные соответственно $\sqrt{v^2(t-t_0)^2+h^2}$ и $\sqrt{[v(t-t_0)-l]^2+h^2}$. (Эту задачу можно решить и чисто геометрически, без дифференциального исчисления.) **7.2.3.** а) $y_{\max}=0$ при $x=0$; б) $y_{\max}=1$ при $x=0$.

7.3.1. Докажите, что $y''<0$. **7.3.2.** а) $\left(\frac{x_1^n+x_2^n}{2}\right)^{1/n}>\frac{x_1+x_2}{2}$ при $n>1$, $x_1\neq x_2$; б) $\left(\frac{x_1^m+x_2^m}{2}\right)^{1/m}<\frac{x_1+x_2}{2}$ при $0<m<1$, $x_1\neq x_2$; в) $1:\left[\frac{1}{2}\left(\frac{1}{x_1^k}+\frac{1}{x_2^k}\right)\right]^{1/k}>\left(\frac{x_1+x_2}{2}\right)$ при $k>0$ и $x_1\neq x_2$ (и x_1, $x_2>0$); г) $^1/_2x_1\log x_1+{}^1/_2x_2\log x_2>{}^1/_2(x_1+x_2)\times$ $\times\log[(x_1+x_2)/2]$ (во всех случаях приведено лишь неравенство, являющееся частным случаем неравенства (1)).

7.4.1. $\pi/2$. **7.4.2.** 1/6. **7.4.3.** $2\pi+4/3$ и $6\pi-4/3$. **7.4.4.** а) $a\ln 2$; б) $\frac{37}{12}a$ (здесь a — количество краски, идущее на окраску единицы площади). **7.4.5.** 10π.

7.6.1. $F(a)+F(b)=\int\limits_1^a\frac{dx}{x}+\int\limits_1^b\frac{dx}{x}=\int\limits_1^a\frac{dx}{x}+\int\limits_a^{ab}\frac{dx}{x}=\int\limits_1^{ab}\frac{dx}{x}=F(ab)$.

7.7.1. $\bar y=\overline{x^2}=\frac{1}{2}\int\limits_0^2 x^2dx=\frac{4}{3}$. При этом $y(1)=1<\bar y\approx 1{,}33<\frac{y(0)+y(2)}{2}=2$.

7.7.2. а) $\bar y=\frac{1}{6}y(0)+\frac{2}{3}y(1)+\frac{1}{6}y(2)=\frac{2}{3}\cdot 1+\frac{1}{6}\cdot 4=\frac{4}{3}$; б) $\int\limits_a^b y\,dx=\left(\frac{1}{3}rx^3+\right.$ $\left.+\frac{1}{2}px^2+qx\right)\Big|_a^b=\frac{1}{3}r(b^3-a^3)+\frac{1}{2}p(b^2-a^2)+q(b-a)=(b-a)\left[\frac{1}{3}r(b^2+ab+a^2)+\right.$ $\left.+\frac{1}{2}p(b+a)+q\right]$. По формуле Симпсона (3) должно быть $\int\limits_a^b y\,dx=(b-a)\,\bar y=$ $=(b-a)\left[\frac{1}{6}y(a)+\frac{2}{3}y\left(\frac{a+b}{2}\right)+\frac{1}{6}y(b)\right]$. Подставляя сюда значения $y(a)$, $y\left(\frac{a+b}{2}\right)$ и $y(b)$ и сравнивая с полученным выше выражением, убеждаемся в том, что они тождественны. **7.7.3.** $\bar F=\frac{1}{2R-R}\int\limits_R^{2R}\frac{A}{r^2}\,dr=\frac{1}{R}\left(-\frac{A}{r}\right)\Big|_R^{2R}=\frac{1}{R}\left(-\frac{A}{2R}+\frac{A}{R}\right)=$ $=0{,}5\frac{A}{R^2}$. Среднее значение силы на этом участке вдвое меньше силы на поверхности Земли, $\bar F=0{,}5F_0$. **7.7.4.** а) $F(R)=F_0$, $F(2R)=\frac{1}{4}F_0$; $\frac{F(R)+F(2R)}{2}=$

$=0,625F_0>0,5F_0$ (ошибка имеет порядок 10%); б) $F\left(\frac{R+2R}{2}\right)=F\left(\frac{3}{2}R\right)=\frac{4}{9}F_0$; поэтому $\frac{1}{6}F_0+\frac{2}{3}\cdot\frac{4}{9}F_0+\frac{1}{6\cdot 4}F_0=\frac{109}{216}F_0\approx 0,505F_0$ (ошибка 1%). **7.7.5.** $\frac{x_0^n}{n+1}$. **7.7.6.** $\frac{m-n}{\ln m-\ln n}$. При $m=n+\nu$, где $\nu\ll n$, $\ln m=\ln(n+\nu)=\ln n+\ln\left(1+\frac{\nu}{n}\right)=$ $=\ln n+\frac{\nu}{n}-\frac{\nu^2}{2n^2}+\ldots$ **7.7.7.** а) Оба средних значения равны $\frac{1}{2}$; б) $\frac{1}{2}-\frac{1}{\pi}$ и $\frac{1}{2}+\frac{1}{\pi}$. **7.7.8.** Если T — период функции y, то должно быть $\sin[\omega(t+T)+\alpha]=$ $=\sin(\omega t+\alpha)$, откуда $\omega T=2\pi$, $T=\frac{2\pi}{\omega}$. Период же функции y^2 равен $\frac{T}{2}=\frac{\pi}{\omega}$; следовательно, надо найти среднее значение функции $y=\sin^2(\omega t+\alpha)$ на промежутке от $t=0$ до $t=\frac{\pi}{\omega}$. Получаем

$$\bar{y}=\int_0^{\pi/\omega}\frac{\sin^2(\omega t+\alpha)\,dt}{\pi/\omega}=\frac{\omega}{\pi}\int_0^{\pi/\omega}\left[\frac{1}{2}-\frac{1}{2}\cos 2(\omega t+\alpha)\right]dt=\frac{1}{2}.$$

7.8.1. а) $s=\int_0^1\sqrt{1+4x^2}\,dx$; б) $s=\int_0^1\sqrt{1+e^{2x}}\,dx$; в) $s=\frac{4}{a}\int_0^a\sqrt{a^2+\frac{b^2x^2}{a^2-x^2}}\,dx$.

7.8.2. Выполнив указанную замену переменной, получим $s=\int_{\sqrt{2}}^{\sqrt{1+e^2}}\frac{z^2}{z^2-1}dz$. Но $\int\left(1+\frac{1}{z^2-1}\right)dz=z+\int\frac{dz}{z^2-1}=z+\int\left(\frac{1/2}{z-1}-\frac{1/2}{z+1}\right)dz=z+\frac{1}{2}\ln\frac{z-1}{z+1}\,(+C)$. Поэтому $s=\left(z+\frac{1}{2}\ln\frac{z-1}{z+1}\right)\Big|_{\sqrt{2}}^{\sqrt{1+e^2}}=\sqrt{1+e^2}-\sqrt{2}+\frac{1}{2}\ln\frac{\sqrt{1+e^2}-1}{\sqrt{1+e^2}+1}-\frac{1}{2}\ln\frac{\sqrt{2}-1}{\sqrt{2}+1}$. **7.8.3.** Разбивая дугу цепной линии на участки от $x=0$ до $x=0,9$ и от $x=0,9$ до $x=2$, находим $s_1\approx 1,043$, $s_2=\frac{e^2+e^{-2}}{2}-\frac{e^{0,9}+e^{-0,9}}{2}+\frac{1}{2}\int_{0,9}^{2}\frac{2}{e^x-e^{-x}}dx$. В последнем интеграле можно положить $e^x=t$, тогда $\int\frac{2dx}{e^x-e^{-x}}=\int\frac{2e^xdx}{(e^x)^2-1}=\int\frac{2dt}{t^2-1}=\int\left(\frac{1}{t-1}-\right.$ $\left.-\frac{1}{t+1}\right)dt$. Окончательно получаем $s_2\approx 2,624$ и $s=s_1+s_2\approx 3,667$. По точной формуле находим $s=3,627$. Ошибка равна примерно 1%. **7.8.4.** $s\approx 1,146$. **7.8.5.** Длина рассматриваемой дуги окружности $S=\int_0^{R/\sqrt{2}}\frac{Rdx}{\sqrt{R^2-x^2}}=\int_0^{R/\sqrt{2}}\left[1+\frac{1}{2}\left(\frac{x}{R}\right)^2+\frac{3}{8}\left(\frac{x}{R}\right)^4+\right.$ $\left.+\frac{5}{16}\left(\frac{x}{R}\right)^6+\frac{35}{128}\left(\frac{x}{R}\right)^8+\ldots\right]dx=R\frac{1}{\sqrt{2}}\left(1+\frac{1}{12}+\frac{3}{160}+\frac{5}{896}+\frac{35}{18\,432}+\ldots\right)$. Соответственно этому получаем следующие оценки числа π: 1) $\pi\approx\frac{4}{\sqrt{2}}\left(1+\frac{1}{12}+\frac{3}{160}\right)\approx$ $\approx 3,117$; 2) $\pi\approx\frac{4}{\sqrt{2}}\left(1+\frac{1}{12}+\frac{3}{160}+\frac{5}{896}\right)\approx 3,133$; 3) $\pi\approx\frac{4}{\sqrt{2}}\left(1+\frac{1}{12}+\frac{3}{160}+\frac{5}{896}+\right.$ $\left.+\frac{35}{18\,432}\right)\approx 3,138$.

7.9.1. а) $k=\frac{ab}{\sqrt{(a^2\sin^2 t+b^2\cos^2 t)^3}}$; б) $k=\frac{2x^3}{\sqrt{(1+x^4)^3}}$; в) $k=\frac{12ax^2}{\sqrt{(1+16a^2x^6)^3}}$.

7.10.2. $V=2\pi$. **7.10.3.** $x_C=\frac{2}{\pi}R$.

7.11.1. а) $y_{\max}=2$ при $x=0$; $y_{\min}=-2$ при $x=2$; б) максимумов и минимумов нет; кривая пересекает ось x в точке $x=1+\sqrt[3]{14}\approx 3,4$, а ось y — в точке $y=-15$;

в) максимумов и минимумов нет; кривая пересекает ось x между точками $x=0$ и $x=-1$, а ось y — в точке $y=3$. **7.11.2.** а) Три корня; б) три корня; в) два корня; г) один корень.

Часть вторая

Глава 8

8.1.1. $T \approx 1600$ лет. **8.1.2.** 176,5 г. **8.1.3.** 53,3 г. **8.1.4.** Мы знаем, что $N(t)=N_0 e^{-t/\bar{t}}$, где N_0 — количество вещества в начальный момент $t=0$. Нас интересует момент времени t_1, к которому осталось $(100-1)\%=99\%$ вещества: $N(t_1)=\frac{99}{100}N_0$. Поэтому $\frac{99}{100}N_0=N_0 e^{-t_1/\bar{t}}$, откуда $t_1=\bar{t}\ln\frac{100}{99}$. Для радия $\bar{t}=2400$ лет, поэтому $t_1=2400\ln\frac{100}{99}=$ $=24$ (года). Аналогично в трех других случаях находим $t_2 \approx 250$ лет; $t_3 \approx 5500$ лет; $t_4 \approx 11\,000$ лет. **8.1.5.** Пусть в начальный момент времени $t=0$ количество атомов в 10^{12} атомов породы равно N_0; в момент $t=10\,000$ это количество равно 1. Поэтому $1=N_0 e^{-\frac{10000}{2400}}$, откуда $N_0 \approx e^4 \approx 65$. Аналогично находим, что 10^6 лет назад $N_0 \approx e^{417} \approx$ $\approx 10^{181}$. Получился явно нелепый результат: в 10^{12} атомов породы содержалось 10^{181} атомов радия! Еще более нелепый результат получается, если подсчитывать содержание радия $5\cdot 10^9$ лет назад. Нелепость результата доказывает неправильность исходного предположения о том, что имеющееся в настоящее время количество радия можно рассматривать как остаток от распада радия, входящего в состав Земли в момент ее образования (ср. с § 8.4).

Глава 9

9.1.1. $A(t)=-h\int\limits_{t_0}^{t} v^2dt$; A отрицательно, так как $\int\limits_{t_0}^{t} v^2dt \geqslant 0$ при $t>t_0$. **9.1.2.** Движение тела периодическое с периодом $T=\frac{2\pi}{\omega}$; требуется определить работу за полпериода. В течение первой четверти периода скорость положительна, а потому $F=-h$; во второй четверти периода скорость отрицательна и $F=h$. В каждом из этих промежутков времени сила постоянна, а работа равна произведению силы на путь, пройденный телом; для 1-й и 2-й четвертей периода $A_1=-hB$ и $A_2=h(-B)=-hB$; работа за полпериода $A=A_1+A_2=-2hB$. **9.1.3.** $A=Bf_0\omega_1\int\limits_0^T \sin\omega_0 t\cos\omega_1 t\,dt$. Этот интеграл легко взять, воспользовавшись тем, что $\sin\omega_0 t\cos\omega_1 t=\frac{1}{2}[\sin(\omega_0+\omega_1)t+$ $+\sin(\omega_0-\omega_1)t]$. Поэтому $A=\frac{Bf_0\omega_1}{2}\int\limits_0^T[\sin(\omega_0+\omega_1)t+\sin(\omega_0-\omega_1)t]\,dt=\frac{Bf_0\omega_1}{2}\times$ $\times\left[\frac{1}{\omega_0+\omega_1}+\frac{1}{\omega_0-\omega_1}-\frac{\cos(\omega_0+\omega_1)T}{\omega_0+\omega_1}-\frac{\cos(\omega_0-\omega_1)T}{\omega_0-\omega_1}\right]$. В случае $\omega_1=\omega_0$ пользоваться последней формулой нельзя. Однако в этом случае $\sin\omega_0 t\cos\omega_1 t=\frac{1}{2}\sin 2\omega_0 t$, откуда $A=\frac{Bf_0\omega_0}{2}\int\limits_0^T \sin 2\omega_0 t\,dt=\frac{Bf_0}{4}(1-\cos\omega_0 T)$. **9.1.4.** Работа силы сопротивления воздуха $A_1(t)=-\frac{aS\rho g^3}{8}t^4$. Работа силы тяжести $A_2(t)=\frac{mg^2}{2}t^2$. Для шарика: $A_1(1)=-0{,}00965$, $A_1(10)=-96{,}5$, $A_1(100)=-965\cdot 10^3$; $A_2(1)=0{,}177$, $A_2(10)=1{,}77$, $A_2(100)=177$. Для пули: $A_1(1)=-1{,}18\cdot 10^{-3}$, $A_1(10)=-11{,}8$, $A_1(100)=-118\cdot 10^3$; $A_2(1)=0{,}435$, $A_2(10)=43{,}5$, $A_2(100)=4350$ (значения A везде в Дж). **9.1.5.** $A=$ $=\frac{aS\rho(v-v_0)^2 b}{2}$. Мощность W определяем, пользуясь формулой $W=Fv$, $W=$ $=\frac{aS\rho(v_0-v)^2 v}{2}$. Найдем, при какой скорости v (при данном v_0) мощность будет

максимальной, т. е. когда $\frac{dW}{dv}=0$; находим $v_1=v_0$ и $v_2=v_0/3$. Ясно, что значение $v=v_0$ нас не интересует, так как оно обращает мощность в нуль. Интересующее нас значение есть $v=v_0/3$ (читатель может произвести полное исследование, основанное на учете знака d^2W/dv^2). При $v_0=30$ м/с, $v=10$ м/с, $W_{max}\approx 25{,}75\cdot 10^5$ Вт.
9.1.6. Работа силы за период $A=B\pi f\sin\alpha$; мощность $W=\frac{Bf\omega}{2}\sin\alpha$.

9.3.1. Прямую, на которой расположены заряды, примем за ось x; пусть заряд e_1 совпадает с началом координат, а заряд e_2 расположен в точке $x=2a$. Равновесие заряда в точке x возможно, лишь если $F=-\frac{du}{dx}=0$. Но $F=\frac{e_1e}{x^2}-\frac{4e_1e}{(2a-x)^2}$, если $0<x<2a$, т. е. если заряд e расположен между зарядами e_1 и e_2; $F=-\frac{e_1e}{x^2}-\frac{4e_1e}{(2a-x)^2}$, если $x<0$, и $F=\frac{e_1e}{x^2}+\frac{4e_1e}{(2a-x)^2}$, если $x>2a$. В первом случае условие $F=0$ дает $x_1=2a/3$, $x_2=-2a$ (второй корень x_2 отбрасываем, так как должно быть $x>0$); в случаях же $x<0$ и $x>2a$ уравнение $F=0$ вообще не имеет решения. Следовательно, имеется одно положение равновесия $x_1=2a/3$. Далее подсчитываем $\frac{d^2u}{dx^2}$ в точке $x=x_1=2a/3$; мы убедимся, что если $e>0$, то $\frac{d^2u}{dx^2}>0$, т. е. равновесие устойчиво; если же $e<0$, то равновесие неустойчиво. 9.3.2. Имеется одно положение равновесия x_1 вне зарядов. Если система координат выбрана так же, как в решении упр. 1, то $x_1=-2a$. Если $ee_1>0$, то равновесие устойчивое, если $ee_1<0$, то равновесие неустойчивое.

9.4.1. Уравнение движения $m\frac{dv}{dt}=F$; пользуясь тем, что $v=0$ при $t=0$, находим $v=\frac{F}{m}t$. Поэтому $\frac{dx}{dt}=\frac{F}{m}t$, $dx=\frac{F}{m}tdt$, откуда $\int_0^x dx=\frac{F}{m}\int_0^t tdt$, так как $x=0$ при $t=0$. Окончательно $x=\frac{F}{2m}t^2$. 9.4.2. а) $x=v_0t+\frac{F}{2m}t^2$; б) $x=x_0+v_0t+\frac{F}{2m}t^2$.
9.4.3. 2,5 м. 9.4.4. Уравнение движения есть $m\frac{dv}{dt}=mg$; отсюда находим $x=\frac{gt^2}{2}$, $t=\sqrt{\frac{2x}{g}}\approx 4{,}5$ с при $x=100$ м. 9.4.5. а) $t=3{,}6$ с. б) $t=5{,}6$ с. Далее, в случае а) $v=gt+v_0$. Пусть $v_п$ — скорость в момент приземления; тогда $v_п=gt_п+v_0$, где $t_п$ — время приземления. Из уравнения $v=gt+v_0$ находим $x=v_0t+\frac{gt^2}{2}$. Пусть шарик падает с высоты H; тогда $x=H$ при $t=t_п$; поэтому $2H=2v_0t_п+gt_п^2$, откуда $t_п=\frac{-v_0+\sqrt{v_0^2+2gH}}{g}$ и $v_п=gt_п+v_0=\sqrt{v_0^2+2gH}$. В случае б) $v=gt-v_0$, остальное аналогично; конечная скорость получается такая же, как и в случае а). 9.4.6. $x=\frac{k}{6m}t^3+v_0t$. 9.4.7. а) $x=-\frac{f}{m\omega^2}\cos\omega t$, $T=\frac{2\pi}{\omega}$, $x_{max}=\frac{f}{m\omega^2}$, $v_{max}=\frac{f}{m\omega}$; б) $x=\frac{f}{m\omega}t-\frac{f}{m\omega^2}\sin\omega t$. 9.4.8. Пусть искомая скорость есть v_0. Тогда закон движения тела $x=x_0+v_0(t-t_0)+\frac{F}{2m}(t-t_0)^2$. Так как $x=x_1$ при $t=t_1$, то $x_1=x_0+v_0(t_1-t_0)+\frac{F}{2m}(t_1-t_0)^2$, откуда $v_0=\frac{x_1-x_0}{t_1-t_0}-\frac{F}{2m}(t_1-t_0)$.

9.8.1. $K=\frac{F^2}{2m}t^2=F(x-x_0)$. 9.8.2. $K=\frac{f^2}{2m\omega^2}\sin^2\omega t$; $K_{max}=\frac{f^2}{2m\omega^2}$. 9.8.3. $\bar K=\frac{mA^2\omega^2}{2}\overline{\sin^2(\omega t+\alpha)}=\frac{mA^2\omega^2}{4}$. 9.8.5. а) $A\approx 4\cdot 10^7$ Дж, $W\approx 2{,}2\cdot 10^5$ Вт; б) $A\approx 7\cdot 10^7$ Дж, $W\approx 38\cdot 10^4$ Вт. 9.8.6. Определим работу каждой из сил в отдельности. Для этого предварительно определим скорость тела. Из уравнения $m\frac{dv}{dt}=at+$

$+a(\theta-t)=a\theta$ находим $v=v_0+\frac{a\theta}{m}t$. Работа силы F_1 есть $A_1=\int\limits_0^\theta at\left(v_0+\frac{a\theta}{m}t\right)dt=$ $=\frac{av_0\theta^2}{2}+\frac{a^2\theta^4}{3m}$. Аналогично находим работу A_2 силы F_2: $A_2=\frac{av_0\theta^2}{2}+\frac{a^2\theta^4}{6m}$. Составим произведение импульса на среднюю скорость: $I_1=\int\limits_0^\theta atdt=\frac{a\theta^2}{2}$, $I_2=$ $=\int\limits_0^\theta a(\theta-t)\,dt=\frac{a\theta^2}{2}$, $\bar v=\frac{\int\limits_0^\theta\left(v_0+\frac{a\theta}{m}t\right)dt}{\theta}=v_0+\frac{a}{2m}\theta^2$, и, значит, $I_1\bar v=\frac{a\theta^2}{2}\left(v_0+\right.$ $\left.+\frac{a}{2m}\theta^2\right)$, $I_2\bar v=\frac{a\theta^2}{2}\left(v_0+\frac{a}{2m}\theta^2\right)$. Видно, что $(I_1+I_2)\bar v=A_1+A_2$, хотя $I_1\bar v\neq A_1$, $I_2\bar v\neq A_2$ (ср. с. 284). 9.8.7. В начале опыта масса m имела скорость v_0 (она двигалась вместе с поездом) и кинетическую энергию $K_1=\frac{mv_0^2}{2}$.

После действия человека скорость массы стала v_0+v_1, $K_2=\frac{m(v_0+v_1)^2}{2}$, где $v_1=\frac{Ft}{m}$. Изменение кинетической энергии $\Delta K=K_2-K_1=m\frac{(v_0+v_1)^2}{2}-\frac{mv_0^2}{2}$. Это есть работа, произведенная над массой поездом и человеком вместе. Для того чтобы определить работу, произведенную над массой m человеком, заметим, что скорость массы по отношению к человеку, едущему в том же поезде, была равна нулю до опыта и стала равной v_1 после опыта. Поэтому работа человека $A_1=\frac{mv_1^2}{2}-0=$ $=\frac{mv_1^2}{2}=\frac{F^2t^2}{2m}$. Теперь легко определить работу A_2 двигателя, толкающего поезд: $A_2=\Delta K-A_1=mv_0v_1=v_0Ft$. (Последний результат легко получить и из других соображений. Действительно, работа двигателя $A_2=\int\limits_0^t vFdt$; так как скорость $v=v_0$ постоянна, то $A_2=v_0\int\limits_0^t Fdt=v_0Ft$.) 9.8.8. До начала опыта скорость массы m и скорость человека равны нулю. После опыта масса m приобрела скорость v_1, человек приобрел скорость v_2. Определим эти скорости из уравнений $m\frac{dv_1}{dt}=F$ и $M\frac{dv_2}{dt}=-F$, поскольку если человек действует на массу m силой F, то масса m действует на человека силой $-F$ (третий закон Ньютона!). Находим $v_1=\frac{Ft}{m}$, $v_2=-\frac{Ft}{M}$. Работа, произведенная силой F над массой m и над человеком, есть $A=K_1+K_2$, где $K_1=$ $=\frac{mv_1^2}{2}=\frac{F^2t^2}{2m}$ — изменение кинетической энергии массы m, $K_2=\frac{Mv_2^2}{2}=\frac{F^2t^2}{2M}$ — изменение кинетической энергии человека. Отсюда $A=\frac{F^2t^2(M+m)}{2Mm}$. 9.8.9. Изменение кинетической энергии массы m равно $\Delta K_m=mv_0v_1+\frac{mv_1^2}{2}$, где $v_1=\frac{F}{m}t$; изменение кинетической энергии человека $\Delta K_M=\frac{Mv_2^2}{2}+Mv_0v_2$, где $v_2=-\frac{F}{M}t$, $A=\frac{F^2t^2(M+m)}{2Mm}$. 9.8.10. а) Если $t_1'=t_1+b$ и $t_2'=t_2+b$, то $\tau'=t_2'-t_1'=t_2-t_1=\tau$. б) Если $x_1'=x_1+$ $+vt_1+a$ и $x_2'=x_2+vt_2+a$, где $t_1=t_2$, то $d'=x_2'-x_1'=x_2-x_1=d$.

9.9.1. Для $\varphi=30^\circ$ имеем $x_2=565$ м, $y_{max}=81{,}5$ м; для $\varphi=45^\circ$ получаем $x_2=$ $=650$ м, $y_{max}=163$ м; для $\varphi=60^\circ$ получаем $x_2=565$ м, $y_{max}=244$ м. 9.9.2. Уравнение траектории таково: $y=x\,\mathrm{tg}\,\varphi-x^2\frac{g}{2v_0^2\cos^2\varphi}$. При заданном $x=500$ м ищем φ, при

котором $y = y_{max}$, т. е. решаем уравнение $\frac{dy}{d\varphi} = 0$. Оно дает $\operatorname{tg}\varphi = \frac{v_0^2}{gx}$, т. е. $\frac{1}{\cos^2\varphi} = \frac{v_0^4 + g^2x^2}{g^2x^2}$. Пользуясь этим, из уравнения траектории находим $y_{max} = \frac{v_0^2}{2g} - \frac{x^2 g}{2v_0^2}$. Полагая $v_0 = 80$ м/с, $x = 500$ м, находим $y_{max} \approx 135$ м.

9.10.1. $r \approx 30\,000$ км.

9.11.1. Найдите вторую производную $\frac{d^2y}{dz^2}$ функции $y = F(z)$ (ср., впрочем, со сказанным на с. 294 и 187—188; заметьте, что $F(z) \to 0$ при $z \to 0$ и при $z \to \infty$).

9.12.1. Совместив начало координат с центром тяжести стержня, получим $I_0 = \int\limits_{-l/2}^{l/2} x^2\sigma dx = \sigma \int\limits_{-l/2}^{l/2} x^2 dx = \sigma \frac{l^3}{12}$; так как $m = \sigma l$, то последний результат можно записать так: $I_0 = m\frac{l^2}{12}$. **9.12.2.** Совместим начало координат с точкой стыковки кусков различной плотности, так что для первого куска плотность равна σ_1 при $x < 0$, а для второго она равна σ_2 при $x > 0$. Тогда $x_c = \frac{1}{2}\frac{\sigma_2 l_2^2 - \sigma_1 l_1^2}{l_1\sigma_1 + l_2\sigma_2}$. **9.12.3.** Выбрав систему координат согласно указанию, получаем $x_c = \frac{2}{3}L$; момент инерции относительно начала $I = \int\limits_0^L x^2\sigma(x)\,dx = \frac{aL^4}{4}$. Так как $m = \frac{aL^2}{2}$, то $a = \frac{2m}{L^2}$; поэтому $I = \frac{mL^2}{2}$. Но в силу (13) $I_0 = I - mL_1^2$, где $L_1 = \frac{2}{3}L$; отсюда $I_0 = \frac{mL^2}{18}$.

9.13.1. а) Ср. с указанием к упр. 12.3. Ответ. Центр тяжести принадлежит медиане треугольника и удален от его вершины на расстояние, равное 2/3 медианы (совпадает с точкой пересечения медиан). б) Пусть основания трапеции равны a и b, где $a > b$; центр тяжести расположен на прямой, соединяющей середины оснований, на расстоянии $\frac{1}{3}h\frac{a+2b}{a+b}$ от нижнего основания, где h — высота трапеции. в) Центр тяжести расположен на оси симметрии полукруга на расстоянии $\frac{4}{3\pi}R$ от ограничивающего полукруг диаметра, где R — радиус полукруга.

9.14.1. Запишем уравнение так: $\frac{dt}{dv} = \frac{1}{\beta(v_1^2 - v^2)}$, откуда, учитывая начальное условие $v(0) = v_0$, находим $t = \frac{1}{\beta}\int\limits_{v_0}^{v}\frac{dv}{v_1^2 - v^2}$. (Для того чтобы выполнить интегрирование, достаточно записать подынтегральную функцию в виде $\frac{1}{v_1^2 - v^2} = \frac{a}{v_1 - v} + \frac{b}{v_1 + v}$ и определить числа a и b методом неопределенных коэффициентов; см. упражнения к § 5.3.) Таким путем получаем $\ln\frac{v_1+v}{v_1-v} = \ln A + 2\beta v_1 t$, где $A = \frac{v_1+v_0}{v_1-v_0}$, или $\frac{v_1+v}{v_1-v} = Ae^{2\beta v_1 t}$, т. е. $v = v_1\frac{Ae^{2\beta v_1 t} - 1}{Ae^{2\beta v_1 t} + 1}$. При очень больших t величина $e^{2\beta v_1 t} \gg 1$, поэтому для таких t будет $v \approx v_1$. Перепишем решение уравнения в виде $v = v_1\frac{A - e^{-2\beta v_1 t}}{A + e^{-2\beta v_1 t}}$, или $v - v_1 = -2v_1\frac{e^{-2\beta v_1 t}}{A + e^{-2\beta v_1 t}}$. При весьма больших t в знаменателе можно пренебречь величиной $e^{-2\beta v_1 t}$, малой по сравнению с A; поэтому в этом случае имеем $v - v_1 \approx -2v_1\frac{e^{-2\beta v_1 t}}{A}$, или $v - v_1 \approx \frac{2v_1(v_0 - v_1)}{v_1 + v_0}e^{-2\beta v_1 t}$; сравнивая это с (24), находим $C = \frac{2v_1(v_0 - v_1)}{v_1 + v_0}$. **9.14.2.** Для случая сопротивления, пропорционального ско-

рости, уравнение движения примет вид $\frac{dv}{dt}=g-\frac{k}{m}v-\frac{A}{m}$, где A — выталкивающая сила. В этом случае происходит установление скорости $v_1=\frac{g}{a}-\frac{A}{ma}$. По закону Архимеда $A=V\rho' g$, где V — объем тела, $\rho'=\rho_ж$ — плотность жидкости. Так как $m=V\rho$ ($\rho=\rho_т$ — плотность тела), то $A=\frac{mg\rho'}{\rho}$ и $v_1=\frac{g}{a}\left(1-\frac{\rho'}{\rho}\right)$. При $\rho'>\rho$ тело всплывает ($v_1<0$), а при $\rho'<\rho$ — погружается ($v_1>0$).

Глава 10

10.1.2. а) Точка остановки $x_0=1/4$. б) Точка остановки $x_0\approx 0{,}9$. в) Точек остановки нет. **10.1.3.** Положение максимума $u(x)$ таково: $u_{max}\approx 9{,}5$ при $x_0=8/3$. Замечая, что левая ветвь графика уходит вверх, а правая — вниз, можем построить очень грубый набросок графика $u(x)$, которого, однако, вполне достаточно для решения упражнения. а) Из постоянства суммы кинетической и потенциальной энергий тела следует, что две точки остановки (в них кинетическая энергия равна нулю) — значения наименьших корней уравнения $-x^3+4x^2=6$. Это уравнение можно решить либо графически, либо при помощи какого-либо численного метода. Получим $x_1\approx$ $\approx -1{,}09$, $x_2\approx 1{,}57$. Тело совершает колебания между точками x_1 и x_2. б) Одна точка остановки $x\approx -2{,}04$. Однако эта точка остановки расположена левее точки, из которой выходит тело в начальный момент времени. Так как начальная скорость направлена вправо, то тело уйдет вправо, не побывав в точке остановки. в) Одна точка остановки $x\approx -2{,}04$. В этом случае тело пройдет влево до точки остановки, а затем уйдет вправо. **10.1.4.** а) Точек остановки нет. Тело уйдет вправо. б) Две точки остановки $x_1=+\sqrt{9/11}$, $x_2=-\sqrt{9/11}$. Движение тела — колебание между точками x_1 и x_2. В случае а) $t=t_0+\int_0^x\sqrt{\frac{1+x^2}{4+3x^2}}\,dx$. В случае б) $t=t_0+\int_{0,5}^x\sqrt{\frac{20+20x^2}{9-11x^2}}\,dx$ при $x<x_1$ и $t=t_1-\int_{x_1}^x\sqrt{\frac{20+20x^2}{9-11x^2}}\,dx$ при $x_1<x<x_2$, где t_1 — время достижения x_1. (Следует заметить, что интегралы остаются конечными, хотя в точках остановки подынтегральная функция обращается в бесконечность.)

10.2.1. а) $x=2\sin t$, б) $x=\cos t$, в) $x=\cos t+2\sin t$. Это решение можно записать в виде $x=C\cos(t+a)$, где $C=\sqrt{5}$, $a=\operatorname{arc\,tg}(-2)\approx -1{,}11$, т. е. $x=\sqrt{5}\cos(t-1{,}11)$. Во всех трех случаях $T=2\pi$.

10.3.1. а) Пусть L — длина маятника. Мы знаем, что $\omega=\sqrt{\frac{mgl}{I}}$, причем в нашем случае $l=\frac{2}{3}L$, $I=\frac{mL^2}{2}$ (см. упр. 9.12.3). Поэтому $\omega=\sqrt{\frac{4g}{3L}}=\sqrt{\frac{8g}{9l}}$, $T=\frac{2\pi}{\omega}=$ $=2\pi\sqrt{\frac{3L}{4g}}\approx 5{,}43\sqrt{\frac{L}{g}}$. Значение l, которому соответствует максимальная частота, определяется по формуле $l_{max}=\sqrt{\frac{I_0}{m}}$. Так как в нашем случае $I_0=\frac{mL^2}{18}$, то $l_{max}=$ $=\frac{L}{\sqrt{18}}$; поэтому $\omega_{max}=\sqrt{\frac{g\sqrt{18}}{2L}}$. Зная ω_{max}, находим $T_{min}=\frac{2\pi}{\omega_{max}}$. б) Здесь $l=\frac{L}{3}$, $I=I_0+ml^2=\frac{mL^2}{18}+\frac{mL^2}{9}=\frac{mL^2}{6}$, $\omega=\sqrt{\frac{2g}{L}}$. Минимальный период здесь совпадает с полученным в случае а), но точка подвеса иная (однако l_{max} — то же, что и раньше).

10.4.1. Пусть колебания задаются формулой $x=C\cos(\omega t+a)$, тогда $v=$ $=-C\omega\sin(\omega t+a)$. Воспользуемся соотношением $kC\frac{dC}{dt}=F_1v$, где $F_1=-hv\,|v|$. $F_1v=-hv^2\,|v|=-hC^3\omega^3\,|\sin^3(\omega t+a)|$, и поэтому закон изменения амплитуды C можно записать так: $kC\frac{dC}{dt}=-hC^3\omega^3A$, где $A=\overline{|\sin^3(\omega t+a)|}$. Заметим, что

$\sin(\omega t+\alpha)$ сохраняет знак при изменении t от $t_1=-\frac{\alpha}{\omega}$ до $t_2=\frac{\pi-\alpha}{\omega}$, причем $t_2-t_1=\frac{T}{2}$. Поэтому $A=\frac{\int\limits_{t_1}^{t_2}\sin^3(\omega t+\alpha)\,dt}{t_2-t_1}=\frac{\omega}{\pi}\int\limits_{-\frac{\alpha}{\omega}}^{\frac{\pi-\alpha}{\omega}}\sin^3(\omega t+\alpha)\,dt$. Заменяя в последнем интеграле переменную $\cos(\omega t+\alpha)=x$, получим $A=-\frac{1}{\pi}\int\limits_1^{-1}(1-x^2)\,dx=\frac{4}{3\pi}$, откуда $\frac{dC}{dt}=-\frac{hC^2\omega^3\cdot 4}{3\pi k}$. Обозначим $\frac{4h\omega^3}{3\pi k}=b$; тогда $\frac{dC}{dt}=-bC^2$ и, следовательно, $\frac{dt}{dC}=-\frac{1}{bC^2}$. Решением этого уравнения является $t=-\frac{1}{b}\frac{C-C_0}{CC_0}$, откуда $C=\frac{C_0}{1+C_0bt}$; здесь C_0 — значение амплитуды в начальный момент времени $t=0$, которое определяется из начальных условий. (Отметим, что такой же закон мы получили для скорости прямолинейного движения в случае сопротивления, пропорционального квадрату скорости (см. (9.14.15)).) **10.4.2.** В этом случае работа за четверть периода равна $-fC$; поэтому средняя мощность есть $-fC\frac{4}{T}=-fC\frac{2\omega}{\pi}$. Получаем уравнение $kC\frac{dC}{dt}=-\frac{2fC\omega}{\pi}$, откуда $\frac{dC}{dt}=-\frac{2f\omega}{k\pi}$ и, значит, $C=C_0-\frac{2f\omega}{k\pi}t$. Колебания прекратятся в момент t_1, когда станет $C=0$, т. е. когда $t_1=\frac{C_0k\pi}{2f\omega}$ (предполагается, что $t_1\gg T$).

10.5.1. Из первого уравнения системы получаем $C_0\cos\varphi=x_0-a$. Пользуясь этим, из второго уравнения нетрудно найти $C_0\sin\varphi=b\frac{\omega}{\omega_1}-\frac{v_0}{\omega_1}-\gamma\frac{x_0-a}{\omega_1}$. Возведя последние два равенства в квадрат и складывая их, получим $C_0^2=\left(b\frac{\omega}{\omega_1}-\frac{v_0}{\omega_1}-\gamma\frac{x_0-a}{\omega_1}\right)^2+(x_0-a)^2$; извлекая из обеих частей равенства квадратный корень, находим C_0; далее $\operatorname{tg}\varphi=\frac{b\frac{\omega}{\omega_1}-\frac{v_0}{\omega_1}-\gamma\frac{x_0-a}{\omega_1}}{x_0-a}$.

10.7.2. Биения (ср. с равенством (5)).

10.8.1. Воспользуйтесь тем, что $\sin^3x=\sin x\sin^2x=\frac{\sin x(1-\cos 2x)}{2}=\frac{\sin x}{2}-\frac{\sin x\cos 2x}{2}=\frac{\sin x}{2}-\frac{\sin 3x-\sin x}{4}=\frac{3}{4}\sin x-\frac{1}{4}\sin 3x$ (представление $f(x)$ в виде (15)). Ответ. $y(x,t)=\frac{3}{4}\cos t\sin x-\frac{1}{4}\cos 3t\sin 3x$. **10.8.2.** Здесь приходится отказаться от условия $D=0$ в (10), так что вместо (14) получаем $y(x,t)=\sum\limits_k b_k\sin\frac{k\pi x}{l}\cos\frac{kc\pi t}{l}+d_k\sin\frac{k\pi x}{l}\sin\frac{kc\pi t}{l}$. **10.8.3.** Условия (2б) дают: $\varphi(x)+\psi(x)=f(x)$; $\varphi'(x)-\psi'(x)=0$, откуда $y(x,t)=\frac{1}{2}[f(x+ct)+f(x-ct)]$, где еще надо доопределить функцию $f(x)$ вне интервала $(0,l)$, объявив ее нечетной и периодической с периодом $2l$.

10.9.3. а) $y=\frac{\pi^2}{3}-4\left(\cos x-\frac{\cos 2x}{2^2}+\frac{\cos 3x}{3^2}+\ldots\right)$; б) $y=\left(\frac{C_1-C_2}{4}\right)\pi-\frac{2(C_1-C_2)}{\pi}\sum\limits_{k=1}^{\infty}\frac{\cos[(2k-1)\pi x/l]}{2k-1}+(C_1+C_2)\sum\limits_{n=1}^{\infty}(-1)^{n+1}\frac{\sin(n\pi x/l)}{n}$.

Глава 11

11.3.1. $p = 1{,}13p_0$; $p = 1{,}48p_0$; $p = 3{,}67p_0$, где p_0 — давление воздуха на поверхности Земли. **11.3.2.** Зависимость давления от высоты имеет вид $p = p_0 e^{-gh/b}$, где $b = 10^3 \frac{RT}{M}$. Для температуры —40° С имеем $T = 273 - 40 = 233$К, $b = \frac{8{,}3 \cdot 10^3 \cdot 233}{29{,}4} \approx 6{,}6 \cdot 10^4$ см²/с²; в этом случае $H = \frac{b}{g} = 6{,}6$ км. Для температуры 40° С имеем $T = 313$ К, $b = 8{,}8 \cdot 10^4$ см²/с², $H = 8{,}8$ км. **11.3.3.** Из уравнения $\frac{dT}{dh} = -\alpha T_0$ следует, что $T = -\alpha T_0 h + C$, где постоянная C определяется из условия: $T = T_0$ при $h = 0$, откуда $C = T_0$; поэтому $T = T_0(1 - \alpha h)$. Основное уравнение для определения плотности таково: $\frac{dp}{dh} = -g\rho$. Воспользуемся уравнением Клапейрона: $p = 10^3 \rho \frac{RT}{M}$; подставляя сюда выражение для T, находим $p = 10^3 \rho \frac{RT_0(1 - \alpha h)}{M}$. Положим $10^3 \frac{RT_0}{M} = b_0$; тогда $p = \rho b_0 (1 - \alpha h)$, откуда $\rho = \frac{p}{b_0(1 - \alpha h)}$. Дифференциальное уравнение для p принимает вид $\frac{dp}{dh} = -g \frac{p}{b_0(1 - \alpha h)}$, или $\frac{dp}{p} = \frac{-g dh}{b_0(1 - \alpha h)}$; интегрируя, получаем $\ln p = \frac{g}{b_0 \alpha} \ln(1 - \alpha h) + C$, откуда $p = A(1 - \alpha h)^{g/b_0\alpha}$, где $A = e^C$. Так как $p = p_0$ при $h = 0$, то $A = p_0$; поэтому $p = p_0(1 - \alpha h)^{g/b_0\alpha}$. **11.3.4.** $p = p_0(1 - 0{,}037 \cdot 10^{-5} h)^{3{,}46}$, $p = 1{,}13p_0$; $p = 1{,}44p_0$; $p = 2{,}97p_0$.

Глава 13

13.2.1. Ток в схеме убывает по закону $j = j_0 e^{-t/RC}$. В интересующий нас момент времени t_1 имеем $j = \frac{9}{10} j_0$; поэтому $\frac{9}{10} j_0 = j_0 e^{-t_1/RC}$, откуда $\frac{9}{10} = e^{-t_1/RC}$. Логарифмируя, находим $\ln \frac{9}{10} = -\frac{t_1}{RC}$, откуда $t_1 = -RC \ln \frac{9}{10} \approx 0{,}105RC$. В силу последней формулы $t_1 \approx 1$ с для $R = 10^7$ Ом; $t_1 \approx 10{,}5$ с для $R = 10^8$ Ом; $t_1 \approx 105$ с для $R = 10^9$ Ом. Момент t_2, когда ток упадет вдвое, определяем аналогично: $0{,}5j_0 = j_0 e^{-t_2/RC}$, откуда $t_2 \approx 0{,}693RC$. Здесь $t_2 \approx 6{,}93$ с при $R = 10^7$ Ом; $t_2 \approx 69{,}3$ с при $R = 10^8$ Ом; $t_2 \approx 693$ с при $R = 10^9$ Ом. **13.2.2.** Воспользовавшись формулой (1.11), получим $\varphi_{C_1} + \varphi_R + \varphi_{C_2} = 0$, откуда $\varphi_R = -(\varphi_{C_1} + \varphi_{C_2})$. Ток в цепи всюду одинаков; поэтому $j = C_1 \frac{d\varphi_{C_1}}{dt} = C_2 \frac{d\varphi_{C_2}}{dt} = \frac{\varphi_R}{R} = -\frac{\varphi_{C_1} + \varphi_{C_2}}{R}$. Получаем уравнения $\frac{d\varphi_{C_1}}{dt} = -\frac{1}{RC_1}(\varphi_{C_1} + \varphi_{C_2})$, $\frac{d\varphi_{C_2}}{dt} = -\frac{1}{RC_2}(\varphi_{C_1} + \varphi_{C_2})$; сложив их, найдем $\frac{d(\varphi_{C_1} + \varphi_{C_2})}{dt} = -\frac{1}{RC} \frac{\varphi_{C_1} + \varphi_{C_2}}{dt}$, где положено $C = \frac{C_1 C_2}{C_1 + C_2}$ (величина C есть суммарная емкость двух последовательно соединенных конденсаторов с емкостями C_1 и C_2). Так как $\varphi_{C_1} + \varphi_{C_2} = a$ при $t = 0$, то из последнего уравнения находим $\varphi_{C_1} + \varphi_{C_2} = ae^{-t/RC}$. Ясно, $C_1 \frac{d\varphi_{C_1}}{dt} - C_2 \frac{d\varphi_{C_2}}{dt} = 0$, или $\frac{d}{dt}(C_1\varphi_{C_1} - C_2\varphi_{C_2}) = 0$. Поэтому $C_1\varphi_{C_1} - C_2\varphi_{C_2} = A$, где A — постоянная. Пользуясь начальным условием: $\varphi_{C_1} = a$, $\varphi_{C_2} = 0$ при $t = 0$, находим $A = C_1 a$. Итак, $\varphi_{C_1} + \varphi_{C_2} = ae^{-t/RC}$, $C_1\varphi_{C_1} - C_2\varphi_{C_2} = C_1 a$. Отсюда находим $\varphi_{C_1} = a \frac{C_1}{C_1 + C_2}\left(1 + \frac{C_2}{C_1} e^{-t/RC}\right)$, $\varphi_{C_2} = a \frac{C_1}{C_1 + C_2}(-1 + e^{-t/RC})$. **13.2.3.** Величины, относящиеся к контуру до увеличения всех линейных размеров, отметим индексом 1, после увеличения размеров — индексом 2. Тогда $T_1 = R_1 C_1$, $T_2 = R_2 C_2$; $C_2 = \frac{\varepsilon S_2}{4\pi d_2} = \frac{\varepsilon n^2 S_1}{4\pi d_1 n} = nC_1$; $R_2 = \rho \frac{l_2}{\sigma_2} = \rho \frac{n l_1}{n^2 \sigma_1} = \frac{R_1}{n}$. Поэтому $T_2 = \frac{R_1}{n} nC_1 = R_1 C_1 = T_1$. Постоянная времени не изменилась.

13.8.1. Пусть разность потенциалов на пластинах конденсатора равна φ. Для схемы рис. 3 $\varphi_E+\varphi_L+\varphi_C=0$, или $L\frac{dj}{dt}+\varphi=E_0$ (ибо $\varphi_E=-E_0$). Так как $j=C\frac{d\varphi}{dt}$, то $LC\frac{d^2\varphi}{dt^2}+\varphi=E_0$, т. е. $LC\frac{d^2\varphi}{dt^2}=-(\varphi-E_0)$, или $LC\frac{d^2z}{dt^2}=-z$, где $z=\varphi-E_0$. Решение последнего уравнения имеет вид $z=B\cos(\omega t+\alpha)$, где $\omega=1/\sqrt{LC}$; поэтому $\varphi=B\cos(\omega t+\alpha)+E_0$. Но $\varphi=0$, $j=0$ при $t=0$; пользуясь этим, находим $B=-E_0$, $\alpha=0$. Окончательно имеем $\varphi=E_0(1-\cos\omega t)$. Значение $\varphi_{max}=2E_0$ отвечает равенству $\cos\omega t=-1$, т. е. $t=\pi/\omega=T/2$; оно достигается через половину периода колебаний. **13.8.2.** Энергия емкости $W=C\varphi^2/2=4CE_0^2/2=2CE_0^2$; энергия, отданная источником напряжения, $P=qE_0=C\varphi E_0=2CE_0^2$.

13.9.1. $j(t)=-\frac{\varphi_0}{L\omega}e^{-\lambda t}\sin\omega t$, где $\lambda=\frac{R}{2L}$, $\omega^2=\frac{1}{LC}-\lambda^2$. Для заданных трех случаев получим: $j(t)=-1{,}0025e^{-0{,}05t}\sin t$; $j(t)=-1{,}031e^{-0{,}25t}\sin 0{,}97t$; $j(t)=-1{,}15e^{-0{,}5t}\sin 0{,}87t$. **13.9.2.** $j(t)=j_0\left(\cos\omega t-\frac{\lambda}{\omega}\sin\omega t\right)e^{-\lambda t}$; формулы для λ и ω см. в упр. 1. Для заданных случаев $j(t)=e^{-0{,}05t}(-\cos t+0{,}05\sin t)$; $j(t)=e^{-0{,}25t}\times(-\cos 0{,}97t+0{,}26\sin 0{,}97t)$; $j(t)=e^{-0{,}5t}(-\cos 0{,}86t+0{,}58\sin 0{,}86t)$. **13.9.3.** Если R велико, то ток через сопротивление мал, т. е. в основном ток идет через индуктивность. Поэтому, чем больше R, тем ближе эта схема к схеме рис. 13.8.1, где $\varphi=\varphi_0\cos(\omega t+\alpha)$. Если R велико, то можно считать, что в схеме рис. 13.9.2 φ имеет такой же вид, но φ_0 есть медленно меняющаяся со временем величина. При этом $\frac{dP}{dt}=-\bar{h}$; но $h=Rj_1^2$, где j_1 — ток, текущий через сопротивление R, $j_1=\frac{\varphi}{R}$; поэтому $h=\frac{\varphi^2}{R}=\frac{\varphi_0^2\cos^2(\omega t+\alpha)}{R}$ и $\bar{h}=\frac{\varphi_0^2}{2R}$. Итак, $\frac{dP}{dt}=-\frac{\varphi_0^2}{2R}$. Замечая, что $P=\frac{C\varphi_0^2}{2}$, находим $\frac{dP}{dt}=C\varphi_0\frac{d\varphi_0}{dt}=-\frac{\varphi_0^2}{2R}$. Отсюда $\frac{d\varphi_0}{dt}=-\frac{1}{2RC}\varphi_0$. Поэтому $\varphi_0=Ae^{-\frac{t}{2RC}}$; $\lambda=\frac{1}{2RC}$.

13.10.1. $\varphi(t)=e^{-t}+te^{-t}$; $\varphi(t)=-0{,}03e^{-5{,}83t}+1{,}03e^{-0{,}17t}$; $\varphi(t)=-0{,}01e^{-9{,}9t}+1{,}01e^{-0{,}1t}$. **13.10.2.** $\varphi(t)=e^{-t}+2te^{-t}$; $\varphi(t)=-0{,}37e^{-3{,}73t}+1{,}37e^{-0{,}27t}$.

Часть третья

Глава 14

14.1.1. а) $u=x^3-3xy^2-3x+1$, $v=3x^2y-y^3-3y$; б) $u=\frac{x^3+x+xy^2}{(1+x^2-y^2)^2+4x^2y^2}$, $v=\frac{y-x^2y-y^3}{(1+x^2-y^2)^2+4x^2y^2}$; в) $u=x^n-C_n^2x^{n-2}y^2+C_n^4x^{n-4}y^4+\ldots$, $v=C_n^1x^{n-1}y-C_n^3x^{n-3}y^3+C_n^5x^{n-5}y^5-\ldots$, где $C_n^k=\frac{n!}{k!(n-k)!}$ — число сочетаний из n элементов по k. **14.1.2.** (6а) и (6б) проверяются исходя из равенств $z^{-1}z=1$ и $(z^{1/3})^3=z$; общая формула (6) вытекает из справедливости ее для любого р а ц и о н а л ь н о г о $n=p/q=(1/q)p$, где отдельно надо рассмотреть случаи положительного и отрицательного p. **14.1.3.** Воспользуйтесь тем, что если $c=r(\cos\alpha+i\sin\alpha)$, то $\sqrt[n]{c}=\rho(\cos\beta+i\sin\beta)$, где $\rho=\sqrt[n]{r}$, и $\beta=\beta_0+\frac{2k\pi}{n}$, где $\beta_0=\frac{\alpha}{n}$, $k=0, 1, 2, \ldots, n-1$. **14.1.4.** В соответствии с результатами упр.3 $\sqrt[4]{-1}=\cos\beta+i\sin\beta$, где $\beta=\pi/4$, $3\pi/4$, $5\pi/4$ $(=-3\pi/4)$, $7\pi/4$ $(=-\pi/4)$; заметьте еще, что во всех случаях величины $\cos\beta$ и $\sin\beta=\pm\sqrt{2}/2$.

14.3.1. Если $\ln u=w$, то $u=e^w$ и $u^z=e^{wz}$; воспользуйтесь этим. **14.3.2.** Ясно, что $e^e\approx(2{,}7)^{2{,}7}>1>i^i$; числа же e^i $(=\cos 1+i\sin 1$, где углы измеряются в радианах) и i^e $(=e^{e\ln i}=e^{(e\pi/2)i})$ несравнимы (по величине можно сравнивать лишь вещественные числа, но не произвольные комплексные числа).

14.4.3. а) Окружность; б) гипербола. Соответственно для а) и для б) $t = \operatorname{tg}\frac{\varphi}{2}$ и $t = \operatorname{th}\frac{\psi}{2}$, если те же линии одновременно заданы и «стандартными» параметрическими уравнениями: а) $x = \cos\varphi$, $y = \sin\varphi$; б) $x = \operatorname{ch}\psi$, $y = \operatorname{sh}\psi$.

Глава 15

15.1.1. Все рассматриваемые производные имеют вид $P\left(\frac{1}{x}\right)e^{-1/x^2}$, где $P(z)$ — многочлен (от переменной z); при $z = 0$ они равны нулю в силу большой скорости убывания функции $y = e^{-1/x^2}$ при $x \to 0$ (ср. § 6.5). Так, например, $f'(x) = 2/x^3 e^{-1/x^2}$; но e^t, где $t = -1/x^2$, убывает при $t \to -\infty$ гораздо быстрее, чем растет при $t \to -\infty$ функция $2|t|^{3/2}$ $(= |2/x^3|)$.

15.2.1. Если $w = \ln z$, то $u = \ln\sqrt{x^2 + y^2}$ и $v = \operatorname{arctg}\frac{y}{x}$, откуда $\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} = \frac{x}{x^2 + y^2}$, $\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x} = \frac{y}{x^2 + y^2}$.

15.2.2. Воспользуйтесь тем, что $u_1 = \frac{\partial u}{\partial x}$, $v_1 = \frac{\partial v}{\partial x}$.

Глава 16

16.2.1. Вблизи точки $x = x_0$ разлагаем функцию $\varphi(x)$ в ряд, сохраняя в нем только первые два члена: $\varphi(x) \approx \varphi(x_0) + \varphi'(x_0)(x - x_0) = \varphi'(x_0)(x - x_0)$. Таким образом, обозначив $\varphi'(x_0) = c$, $x - x_0 = y$, мы получим $\delta(\varphi(x)) = \delta(cy) = \frac{1}{|c|}\delta(y) = \frac{1}{|\varphi'(x_0)|}\delta(x - x_0)$. В более общем случае функция $\delta(\psi(x))$ представляет собой сумму подобных выражений, распространенных по всем нулям функции ψ. **16.2.2.** Функция $\varphi(x) = \sin x$ обращается в нуль при $x_0 = k\pi$, $k = 0, \pm 1, \pm 2, \ldots$, $|\varphi'(x_0)| = |\cos x_0| = |\cos k\pi| = 1$; поэтому $\delta(\sin x) = \sum_{k=-\infty}^{+\infty}\delta(x - k\pi)$ и $\int_{-\infty}^{+\infty}\psi(x)\,\delta(\sin x)\,dx = \sum_{k=-\infty}^{+\infty}\psi(k\pi)$.

16.3.1. а) $y'(x) = 1 - \delta(x - 1)$. б) Так как $y(+0) = 1$, $y(-0) = 0$, то скачок Δy переменной y при $x = 0$ таков: $\Delta y = 1$. Далее, при $x \neq 0$ имеем $y'(x) = -\frac{e^{1/x}}{x^2(1 + e^{1/x})^2}$; поэтому окончательно $y'(x) = \delta(x) - \frac{e^{1/x}}{x^2(1 + e^{1/x})^2}$.

16.4.1. Умножьте правую часть (8) на $\sin\frac{x}{2}$ и воспользуйтесь тем, что $\cos kx \sin\frac{x}{2} = \frac{1}{2}\left[\sin\left(k + \frac{1}{2}\right)x - \sin\left(k - \frac{1}{2}\right)x\right]$.

Глава 17

17.3.1. а) Если Γ — линия тока, то в силу определения функции ψ касательный вектор $\mathbf{t} = \left(-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}\right)$ к линии Γ (т. е. к линии $\psi = \mathrm{const}$) совпадает с вектором $\mathbf{v}$ скорости потока, откуда уже и следует, что вдоль линии движения жидкости, имеющей в каждой точке направление $\mathbf{v} = \mathbf{v}(x, y)$, приращение $d\psi$ функции тока равно нулю. б) Пусть Γ — малая дуга с концами $P = P(x, y)$, $Q = Q(x + dx, y + dy)$, а $R = R(x + dx, y)$; рассмотрите поток жидкости через ломаную PQR, принимая вектор скорости жидкости равным $\mathbf{v}(x, y)$, и докажите, что (с точностью до бесконечно малых 1-го порядка) этот поток равен $d\psi = \psi(x + dx, y + dy) - \psi(x, y)$. **17.3.2.** Воспользуйтесь тем, что векторы касательных к линиям $\varphi = \mathrm{const}$ и $\psi = \mathrm{const}$ в точке $M(x, y)$ таковы: $\mathbf{v}_\varphi = \left(-\frac{\partial\varphi}{\partial y}, \frac{\partial\varphi}{\partial x}\right)$ и $\mathbf{v}_\psi = \left(-\frac{\partial\psi}{\partial y}, \frac{\partial\psi}{\partial x}\right)$, т. е. $\mathbf{v}_\varphi = (v_y, -v_x)$ и $\mathbf{v}_\psi = (v_x, v_y)$, далее воспользуйтесь условием перпендикулярности двух векторов. **17.3.4.** а) При a вещественном линии тока суть прямые, параллельные оси Ox; эквипотенциальные линии параллельны оси Oy; случай чисто мнимого a дает сопряженное

описанному движение жидкости. б) При a вещественном линии тока — гиперболы $xy=\text{const}$; эквипотенциальные линии — гиперболы $x^2-y^2=\text{const}$; величина скорости в точке M пропорциональна расстоянию OM. в) При a вещественном линии тока — окружности, касающиеся в точке O оси Ox; эквипотенциальные линии — окружности, касающиеся в точке O оси Oy. г) Для $w=\ln z$ линии тока — исходящие из O прямые; эквипотенциальные линии — окружности с центром O.

17.4.1. Приравнивая нулю (равновесие!) сумму проекций сил на направление оси y, получим для случая малых отклонений ($y \ll l$): $1-k\dfrac{y_1}{x_1}-k\dfrac{y_1}{l-x_1}=0$. Отсюда

$$y_1(x_1)=1\Big/\left(\frac{k}{x_1}+\frac{k}{l-x_1}\right)=x_1\frac{(l-x_1)}{kl},$$

$$y(x,\ x_1)=\begin{cases}\left(\dfrac{x}{x_1}\right)\Big/\left(\dfrac{k}{x_1}+\dfrac{k}{l-x_1}\right)=\dfrac{x(l-x_1)}{kl}, & 0\leqslant x\leqslant x_1,\\[2ex] \left(\dfrac{l-x}{l-x_1}\right)\Big/\left(\dfrac{k}{x_1}+\dfrac{k}{l-x_1}\right)=\dfrac{x_1(l-x)}{kl}, & x_1\leqslant x\leqslant l.\end{cases}$$

Эта функция $y(x,\ x_1)$ и называется *функцией Грина* задачи о струне. При произвольно распределенной силе $f(x)$ отклонение струны дается формулой $y(x)=$ $=\int\limits_0^l f(x_1)\,y(x,\ x_1)\,dx_1$. Отметим, что с помощью функции Грина мы нашли решение для функции $y(x)$, даже не зная, какому уравнению она подчиняется (это уравнение имеет вид $\dfrac{d^2y}{dx^2}=-\dfrac{f(x)}{k}$, $y(0)=0$, $y(l)=0$). **17.4.2.** Общее решение уравнения без вынуждающей силы есть $x(t)=C_1\sin\omega t+C_2\cos\omega t$, $\omega=\sqrt{k/m}$, C_1 и C_2 — произвольные константы. Так как δ-функция отлична от нуля только при $t=\tau$, то решение уравнения с δ-образной силой и состоянием покоя при $t=-\infty$ имеет вид

$$x(t)=\begin{cases}0 & \text{при } -\infty<t<\tau,\\ C_1\sin\omega t+C_2\cos\omega t & \text{при } \tau<t<+\infty;\end{cases}\qquad (*)$$

δ-образная сила сообщает телу единичный импульс, поэтому покоящееся тело после действия δ-силы приобретает начальную скорость $v_0=\dfrac{\Delta p}{m}=\dfrac{1}{m}$, начальное положение остается равным нулю. Решение уравнения колебаний с такими начальными условиями в момент $t=\tau$ имеет вид

$$x(t,\ \tau)=\begin{cases}0 & \text{при } -\infty<t<\tau,\\ \dfrac{1}{m\omega}\sin\omega(t-\tau) & \text{при } \tau<t<+\infty.\end{cases}$$

Иначе говоря, в формуле (*) $C_1=\dfrac{\cos\omega\tau}{m\omega}$, $C_2=-\dfrac{\sin\omega\tau}{m\omega}$.

Решение задачи с п р о и з в о л ь н о й силой $f(t)$ дается формулой

$$x(t)=\int\limits_{-\infty}^{+\infty} f(\tau)\,x(t,\ \tau)\,d\tau=\int\limits_{-\infty}^{t} f(\tau)\,\frac{1}{m\omega}\sin[\omega(t-\tau)]\,d\tau.$$

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

ЧАСТЬ ВТОРАЯ

ПРИЛОЖЕНИЯ ВЫСШЕЙ МАТЕМАТИКИ К НЕКОТОРЫМ ВОПРОСАМ ФИЗИКИ И ТЕХНИКИ

ЧАСТЬ ТРЕТЬЯ

ДОПОЛНИТЕЛЬНЫЕ ТЕМЫ ИЗ ВЫСШЕЙ МАТЕМАТИКИ

Яков Борисович
ЗЕЛЬДОВИЧ,

Исаак Моисеевич
ЯГЛОМ

ВЫСШАЯ МАТЕМАТИКА ДЛЯ НАЧИНАЮЩИХ ФИЗИКОВ И ТЕХНИКОВ

Утверждено к печати
Институтом теоретической физики
им. Л. Д. Ландау АН СССР

Редактор Л. И. Головина
Редактор издательства Л. Е. Кононенко
Художник В. Н. Типунов
Художественный редактор Т. П. Поленова
Технические редакторы
Т. Д. Панасюк, Л. Н. Золотухина
Корректоры
Н. Б. Габасова, Г. Н. Лащ, Л. В. Лукичева

ИБ № 21107

Сдано в набор 28.12.81
Подписано к печати 06.07.82. Т-09882
Формат $70 \times 108^1/_{16}$
Бумага книжно-журнальная
Гарнитура обыкновенная
Печать высокая
Усл. печ. л. 44,8. Усл. кр. отт. 44,8. Уч.-изд. л. 47,6
Тираж 30000 экз. Тип. зак. 1065
Цена 3 р. 20 к.

Издательство «Наука»
117864 ГСП-7, Москва, В-485, Профсоюзная ул., 90

Ордена Трудового Красного Знамени
Первая типография издательства «Наука»
199034, Ленинград, В-34, 9 линия, 12